ESSAYS IN APPLIED ECONOMICS

by

Marco Carlo Cornia

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

December, 2015

Abstract

This dissertation is composed of three chapters. In the first chapter, I estimate the effect of China's reserve intervention on the Renminbi-Dollar exchange rate and the trade balance in a partially identified vector autoregression. Exogenous reserve intervention is identified by exploiting the exchange rate regime shift of July 21st, 2005. In particular, I exploit the fact that policy is entirely passive in the fixed exchange rate regime, while it is not in the managed float. I analyze the impact of public saving, via reserve intervention, on the economy's net external position to determine the contribution of policy to global imbalances. My estimates indicate that a surprise purchase of 24.2 Billion USDs depreciates the bilateral Renminbi-Dollar rate by about 1% on impact and that this depreciation is highly persistent. The effect on the trade balance is statistically insignificant and economically small in contrast to what is postulated by a growing recent theoretical literature.

In the second chapter, my co-author and I document, using a newly constructed historical dataset, gross bank credit flows for the state of New York between 1912 and 1932. We focus our analysis on this historical period due to the unit banking structure of the financial sector. The lack of branching allows us to measure heterogeneity in credit expansion and contraction across banks to a finer degree. We find higher levels of gross credit creation, destruction and excess reallocation than those documented in the literature. In addition, we document how the cyclical properties and relative volatility of credit creation and destruction vary across collateralized and uncollateralized lending.

In the third chapter, my co-authors and I provide empirical evidence documenting how price dispersion moves with the business cycle in the airline industry. Performing a fixed-effects panel analysis on seventeen years of data covering two business cycles, we find that price dispersion is highly pro-cyclical. This effect is especially pronounced for legacy carriers relative to low-cost carriers. We show that our empirical result is consistent with firms implementing second-degree price-discrimination tactics.

Advisor: Jonathan Wright

Acknowledgments

I am sincerely grateful to my advisers Jonathan Wright, Olivier Jeanne and Jon Faust for their insights, support and guidance throughout this project. I also thank Michele Mazzoleni, Emek Karaca, Burcin Kisacikoglu, Haelim Park and Benjamin Kay for helpful comments on the first chapter of this dissertation as well as seminar participants at the Johns Hopkins University and Office of Financial Research.

I am also thankful to my friends for making my stay in the U.S. a fun and productive one. My family provided much support during my studies. Finally, I am immensely grateful to my girlfriend, Judith Trasancos Rodriguez. Without her immense patience, dedication, and support during all these years, none of this would have been possible.

Contents

1 Global Imbalances and Reserve Intervention: Identifying the Impact of Chinese Exchange Rate Policy 1

1.1 Introduction 1

1.2 Empirical approach 6

1.2.1 VAR specification 6

1.2.2 Exchange rate regimes in China 9

1.3 Identification 11

1.3.1 The VAR Identification Problem 11

1.3.2 Identification of the policy shocks 13

1.3.3 Identifying assumptions 17

1.3.4 Relationship to the literature 22

1.4 Estimation and Inference 23

1.5 Measuring FX intervention 26

1.6 Empirical Results 30

1.6.1 Robustness checks 32

1.7 Support for the identifying assumptions 35

1.7.1 Identification of the policy shocks 35

1.7.2 Data on Non-Deliverable Forwards 37

1.7.3 Outlier analysis and results 37

1.8 Conclusion 39

1.9 Appendix 41

2 Gross Bank Credit Flows: Evidence from the unit banking system in the state of

New York, 1912-1932 - with Haelim Park **52**

2.1 Introduction 52

2.2 Data 55

2.3 Methodology 57

2.4 Gross Credit Flows 62

2.4.1 Creation, Destruction and Reallocation 62

2.4.2 Gross credit flows and collateral 68

2.4.3 Heterogeneous regional shocks 72

2.4.4 Robustness checks 74

2.5 Conclusion 79

2.6 Appendix 81

3 Price dispersion over the business cycle: Evidence from the airline industry - with Kristopher S. Gerardi and Adam H. Shapiro **91**

3.1 Introduction 91

3.2 Data 93

3.2.1 Operating Cost 95

3.3 Estimation 96

3.3.1 Results 98

3.4 Discussion 100

3.4.1 Second-Degree Price Discrimination 100

3.4.2 Stochastic-Demand Pricing 105

3.5 Conclusion 109

3.6 Appendix 110

Bibliography **123**

4 Curriculum Vitae **129**

List of Tables

1.1 Priors and Posteriors for currency shares 50

1.2 Recapitalizations undertaken by the PBOC 50

1.3 High-Frequency based estimates of RMB-USD IRF 51

2.1 Evolution of the State Chartered Banks and Trusts in NY State, 1912-1932 . . 86

2.2 Creation, Destruction and Excess Reallocation 87

2.3 Creation, Destruction and Excess Reallocation: Correlation with the Business Cycle 87

2.4 Creation, Destruction and Excess Reallocation by collateral type 88

2.5 Creation, Destruction and Excess Reallocation by collateral type: Correlation with the Business Cycle 88

2.6 Spurious Credit Destruction: Probit Analysis of exiting banks 89

2.7 Spurious Credit Creation: Probit Analysis of entering banks 89

2.8 Creation, Destruction and Excess Reallocation in NYS 90

3.1 Summary Statistics 118

3.2 Full Sample Estimates 118

3.3 Full Sample Estimates: Percentiles 119

3.4 Panel Estimates by Carrier Type 119

3.5 Panel Estimates by Route Type 120

3.6 Estimates with Capacity Utilization Control: IQR 121

3.7 Estimates with Capacity Utilization Control: Gini 122

List of Figures

1.1 RMB-USD exchange rate regimes 44
1.2 Estimate of FX intervention time-series 45
1.3 Impulse Reponse Functions: Baseline results 46
1.4 Impulse Reponse Functions: Excluding October 2008-July 2010 47
1.5 Impulse Reponse Functions: Diffuse prior 48
1.6 Changes in the spot and forward RMB-USD rates in the event windows . . 49

2.1 Assets and Liabilities Reported 81
2.2 Net credit growth and excess reallocation 82
2.3 Business Cycle proxies: National and State level 83
2.4 Creation and Destruction in NY state 84

3.1 Price Dispersion over the Business Cycle 114
3.2 Pricing Dynamics 115
3.3 Aircraft Operating Costs 116
3.4 Aircraft Utilization Rate 117

CHAPTER 1

Global Imbalances and Reserve Intervention: Identifying the Impact of Chinese Exchange Rate Policy

1.1 Introduction

A striking feature of the last two decades has been the parallel rise of global imbalances and the rapid accumulation of foreign exchange (FX) assets in a number of emerging economies. China, in particular, has accumulated FX assets[1] at an impressive rate in this period and largely in parallel to persistent current account surpluses. The rationale behind this saving behavior of emerging economies has been a hotly debated issue in the profession and policy circles (Bernanke (2005)). Part of the literature emphasizes structural characteristics (Carroll and Jeanne (2009), Caballero and Gourinchas (2008)) intrinsic to these surplus economies that induce agents to save. The unifying feature of this literature is that the saving rate is determined by the behavior of the private sector. The fact that the vast majority of these savings are channeled via the formal sector into FX reserves is happenstance in most of these models (Mendoza and Rios-Rull (2007) being an exception). A diametrically opposite explanation, often referred to as "neo mercantilism" (Dooley and Garber (2003))[2], argues instead that the accumulation of FX reserves is the thrust of global imbalances. In particular, this theory suggests that it is public sector saving, in the form of FX accumulation by emerging market Central Banks, that is the driving force of the observed saving patterns. It is therefore in these policy decisions and not in private sector saving behavior that global imbalances have their origin.

[1] In a slight abuse of terminology I use throughout the paper the terms FX intervention, FX accumulation, reserve intervention, reserve accumulation and FX asset purchases interchangeably.

[2] A third explanation for this phenomenon attributes the rise of global imbalances to structural changes occurring in the deficit countries. See for example Engel and Rogers (2006)

This paper seeks to contribute to this literature by empirically measuring the effectiveness of FX intervention in inducing improvements in the trade balance. The relevance of the "neo mercantilistic" hypothesis is inherently tied to this parameter. In fact, if the effect of FX intervention on the trade balance is small then policy cannot be the main determinant of the observed saving patterns. To do this, I identify exogenous purchases of FX assets by the People's Bank of China (PBoC) and study the dynamic response of the bilateral RMB-USD spot rate and the trade balance to such a shock. I investigate this causal mechanism using China as a case study for essentially four reasons: (i) China is the largest holder of FX reserves worldwide and has built up these reserves largely in parallel to a string of trade surpluses, (ii) the PBoC uses FX purchases in the onshore FX interbank market to guide the bilateral onshore RMB-USD exchange rate (iii) all models that emphasize the importance of this causal channel require a form of friction between domestic and foreign currency bonds. China naturally fits this requirement since its capital control measures limit the domestic sector's ability to access international financial markets (Jeanne (2012), Bacchetta et al. (2013)), (iv) the policy relevance of China in the debate on global imbalances and FX accumulation.

To my knowledge, this is one of the first papers that attempts to empirically measure this causal link between FX intervention and the trade balance. Other papers (Aizenman and Lee (2007), Cheung and Ito (2009)) have used panel and cross-sectional methods to measure the importance of FX accumulation on the current account and/or trade balance over the medium run. However, their regression estimates confound the direct effect of FX intervention on the trade balance with feedback effects arising from the endogenous policy reaction of the Central Bank. Thus in the same way that regressing the federal funds rate against output growth cannot inform us about the effect of monetary policy surprises, a regression of FX intervention on the trade balance cannot identify the effect of FX intervention. With the exception of Bayoumi et al. (2015), therefore, this previous literature[3] mea-

[3] In fairness, part of these papers were not written with the intention to tease out causality from the reduced form relationship, but rather to document the correlates of the current account and/or trade balance in the medium run.

sures reduced form correlations while my work attempts to quantify a casual mechanism. In addition to the effect of FX intervention on the trade balance, my empirical approach also estimates the effect of FX intervention on the exchange rate. There is a large empirical literature that attempts to estimate this parameter (Sarno and Taylor (2001)). However, this literature identifies the effect of FX intervention in open economies with a free floating exchange rate. I identify this parameter in a country with a controlled exchange rate regime and closed capital account. Thus, the resulting estimates have fundamentally different interpretations.

The structural VAR approach I follow explicitly allows for there to be simultaneous determination between the trade balance, FX intervention and the exchange rate. Identification of the model is achieved by exploiting the shift in the exchange rate regime that occurred on July 21st, 2005. In particular, identification is achieved as follows. First, I estimate the conditional variance of FX intervention in the fixed exchange rate regime. In this period, aggregate USD inflows, capital controls and restrictions on domestic holding of foreign assets forced the PBoC to passively absorb the excess supply of USDs onshore via FX intervention. In this policy regime, FX intervention is only endogenously determined, as the PBoC only intervenes to implement the peg to the USD. Thus this reduced form variance identifies variation in FX intervention that is driven by fundamental shocks other than the policy shock. Second, the main identifying assumption is that this variance is constant across exchange rate regimes. Third, in the managed float regime, FX intervention is determined by both endogenous policy - as before - and exogenous FX purchases driven by policy shocks. Knowledge of the magnitude of this endogenous variation allows me to partition total FX variation into endogenous and exogenous components and identify the parameters of the policy equation. Fourth, subject to the assumption that the Central Bank acts to stabilize the RMB-USD rate in response to fundamental shocks (i.e. "leans against the wind"), I show that the policy function parameters are globally identified. Intuitively, the identification strategy recovers the direct effect of FX intervention by expressing priors over the sign and size of the feedback effects that confound the inference.

To preview my results, I find that in response to an exogenous purchase by the PBoC of 24.2 Billion USD the bilateral RMB-USD rate will depreciate by 1% on impact. The depreciation is highly persistent and is statistically significant for at least 2 years after the initial shock. The effect on the trade balance, however, is economically small (-97.2 Million USDs). At the posterior mode, this effect is economically small not only on impact but as well for the whole first two years after the policy shock. In terms of the per dollar effectiveness of FX intervention on the trade balance, at the mode the impact effect is approximately zero and it remains of economically small magnitudes (less than 0.1 per dollar of FX intervention) even at longer impact horizons. Looking at the distribution of this parameter, in the 68% highest likelihood set of models, the largest positive impact effect estimates that the trade balance improves only by ten cents (0.1 USD) for every dollar of FX intervention. When I compute this same measure over longer horizons I find similar results. In particular, I compute the effectiveness of FX intervention on the trade balance over a 2 year period successive to the initial policy shock. In the 68% highest likelihood set of models, the largest effect I find is an increase of about 0.1 dollars in the trade balance for every dollar of FX intervention. These results suggest that surprise purchases of FX affect the onshore RMB-USD rate in the expected direction but do not have any economically significant effects on the trade balance. Though my estimates only identify the effect of surprise FX intervention, the results I have illustrated do not support the current formulation of the "neo mercantilistic" hypothesis.

The results outlined above also have important implications for international policy coordination. In fact, under the neo-mercantilistic hypothesis, the policy maker may use a combination of foreign asset purchases and capital controls to influence its external balance. This is effectively a form of untargeted export subsidy (Korinek and Serven (2010), Jeanne (2012)) and the trade distortions that it induces are forbidden under World Trade Organization regulations. From this point of view then, the case for common rules on capital account policies seems no different than that on trade policies. My estimates inform this debate by showing that, in fact, such policies do not seem to be effective in this dimension. Further, though current account imbalances have been largely tempered as a consequence of the financial crisis, the explicit pursuit of FX accumulation policies by emerging market economies appears to be by no means over. Dominguez et al. (2012) document that while reserve accumulation has considerably slowed or even reversed during the crisis, from 2011 onward there has been a return to pre-crisis trends. Thus, the decade-long upsurge in global reserve assets seems likely to endure in this coming decade and understanding the impact and spillovers of such FX intervention policies remains of paramount importance.

Literature: This paper is related to a number of literatures. First, this paper is related to the literature on FX intervention in emerging market economies. It is, on the one hand, related to the empirical work of Disyatat and Galati (2007) and Adler and Tovar (2011) on the role reserves play in resisting appreciation and capital inflows. On the other hand, it is related to the theoretical literature on sterilized FX accumulation that clarifies under what assumptions public flows can affect private consumption-saving decisions in a small open economy. This work stresses the importance of imperfect capital mobility and studies the effect of reserve policy on consumption, exchange rates and interest rates (Aguair and Amador (2011), Benigno and Fornaro (2012), Jeanne (2012), Kumhof (2010)).

This paper is also related to the mercantilistic hypothesis of Dooley and Garber (2003), the related literature on exchange rate undervaluation (Rodrik (2008)) and the precautionary savings approach to reserves (Jeanne and Ranciere (2011)). In the first literature, reserve accumulation emerges as the outcome of a policy directed to resist currency appreciation and promote export led growth. Korinek and Serven (2010) analyze such an economy with a learning-by-investing externality and characterize its welfare properties. The authors find that the social planner's optimal policy leads to both higher consumption growth and faster output growth in equilibrium. Rodrik (2008) similarly documents that countries with depreciated real exchange rates achieve faster growth in the data. The combination of these literatures provides the rationale for active reserve policy. The first explains what possible externalities the policy maker internalizes by pursuing active policy; the second argues that the precautionary paradigm cannot explain the reserve build up of this decade.

This work is also connected to the literature on global imbalances. This set of papers explains current account surpluses in 'catch-up' countries through aggregate risk (Jeanne (2007), Jeanne and Ranciere (2011)), idiosyncratic risk (Carroll and Jeanne (2009)) and financial market imperfections (Caballero and Gourinchas (2008)) or a combination of them (Mendoza and Rios-Rull (2007)). Differently from the explanation emphasized in this paper, in these models, reserve accumulation emerges as a consequence of the private sector's saving decision as these savings are channeled through the official sector into reserves. In this explanation, the role of policy, as emphasized above, is completely absent. Furthermore, these are explanations that characterize the long run behavior of economies rather than short-medium run fluctuations implicitly studied in this paper.

The remainder of the paper is organized as follows. Section 1.2 describes the empirical model. Section 1.3 illustrates the identification argument and the assumptions. Section 1.4 describes the estimation strategy. Section 1.5 illustrates my measure of FX intervention. Section 1.6 discusses the empirical results. Section 1.7 discusses supporting evidence for my identifying assumptions. Section 1.8 makes concluding remarks.

1.2 Empirical approach

In this section, I first describe the structure of FX markets in China to help motivate my VAR specification. Second, I describe the behaviour of the RMB-USD exchange rate in my sample period.

1.2.1 VAR specification

FX markets in China are structured as a two-tier system comprised of a retail market and an interbank market[4]. In the lower tier of the system, the retail market, firms and individuals demand and supply FX exclusively to a set of designated banks. These designated FX banks and the PBoC form the upper tier of the system that is the onshore FX interbank market. The crucial characteristic of this onshore FX market is that existing regulations

[4]The regulations that separate these two trading venues have been gradually changing. For example, since August 2005 non-financial firms that earn and spend FX may directly trade in the interbank market. For details see Jikang and Yuanyuan (2006).

segment it from offshore demand and supply conditions[5]. Thus, if under the prevailing onshore spot rate a net excess supply (demand) of USDs arises in the interbank market, there will be pressure for the onshore RMB-USD rate to appreciate (depreciate).

Over the period in question, China has been subject to large net inflows of USDs arising from a combination of pronounced and persistent current account surpluses and net capital inflows. Due to restrictions on FX holdings by domestic entities, such as the compulsory FX settlement system[6], these USD inflows percolate into the onshore FX interbank market as firms and investors surrender USDs to the designated FX banks in exchange for RMBs. In turn, these banks use the interbank market to square off their FX positions derived from these retail trades. This combination of restrictions on FX holdings and the FX market structure imply that the consolidated net supply (demand) of USDs in the onshore interbank market is inherently connected to China's net external position. Due to the surpluses recorded during the period in question, FX banks as a whole consistently supplied USDs to the onshore interbank market.

When such an imbalance occurs, the only source of demand for this onshore supply of USDs is ultimately the PBoC. Hence, due to its predominant demand role in this market, the PBoC uses FX intervention to curb the excess USD supply in the interbank market and influence the equilibrium onshore rate. These USD purchases are then sterilized through the issuance of Central Bank bills in order to control the money supply (Aizenman and Glick (2009), Ouyang et al. (2010), Zhang (2010). Therefore, throughout the paper, I will use the term FX intervention to signify sterilized FX intervention.

In light of the discussion above, consider the following demand and supply model of the

[5]For more details on the various forms of interaction between onshore and offshore markets, see The offshore Renminbi, by HSBC Global Research

[6]Under the FX purchase and sale system, exporters and foreign investors must surrender at least 75% of their FX earning to designated FX banks. These banks must, in turn, sell their foreign exchange on the interbank FX market.

onshore FX interbank market[7]:

$$\alpha_1 I_t = s_t - s_{t|t-1} - \alpha_2 TB_t - \epsilon_t^D \tag{1.1}$$

$$I_t = \beta_0 + \beta_1 s_t + \beta_2 TB_t + \epsilon_{1,t} \tag{1.2}$$

$$TB_t = \gamma_0 + \gamma_1 s_t + \gamma_2 I_t + \epsilon_{2,t} \tag{1.3}$$

where I_t is the level of FX purchases[8] measured in USDs undertaken by the PBoC during month t, TB_t is the trade surplus (deficit) at month t measured in USDs and s_t is the log of the onshore RMB-USD spot rate measured on the last day of the month. Finally, $\bar{s}_{t|t-1}$ is the PBoC's exchange rate target. The spot rate and the policy target are expressed in local currency units per dollar. Therefore, an increase (decrease) reflects a depreciation (appreciation) of the RMB. Positive (negative) FX interventions measure purchases (sales) of FX assets by the PBoC. Positive (negative) values for TB_t are surpluses (deficits).

Equation (1.1) is the policy reaction function of the PBoC or the demand equation with associated policy shock ϵ_t^D. In each exchange rate regime, the PBoC pursues a policy path $\bar{s}_{t|t-1}$ that it implements using FX interventions to influence the onshore RMB-USD rate, s_t. In particular, in response to appreciation (depreciation) pressure the PBoC intervenes in the interbank market by buying (selling) USDs. These dollars are successively reinvested in FX assets which, in fact, form a sizeable component of the asset side of the PBoC's balance sheet. The extent to which the PBoC reacts to appreciation/depreciation pressures is measured by the parameter α_1. As the value of this parameter increases in absolute value, the policy maker's demand for FX becomes more elastic and the exchange rate becomes more "market determined". On the other hand, at $\alpha_1 = 0$, demand for FX is perfectly inelastic and the policy maker fully determines the exchange rate. In the terminology of the literature, this parameter captures the degree to which Central Banks lean against the wind.

[7] Here, I ignore any lags that may be needed in the specification of this model for brevity. The identification problem and approach remain unchanged with the inclusion of lags in the system formed by (1.1), (1.2), (1.3).

[8] The construction of the FX intervention variable is discussed in detail in Section 1.5

The policy path - $\bar{s}_{t|t-1}$ - captures the broad trend of the exchange rate regime and is assumed to be, at least, contemporaneously exogenous. In the fixed exchange rate period this holds trivially as it is constant. After the July 2005 revaluation, it captures the predictable appreciation trend followed by the PBoC[9]. I also assume that the PBoC sets this policy path in terms of the bilateral RMB-USD rate. This is a natural assumption to make. In fact, Chinese authorities have used the USD as a benchmark to value the RMB at least since 1994, and held a *de facto* peg against the dollar from 1997 to July 2005. After the July 2005 revaluation, they have explicitly managed the appreciation against the dollar using daily central parity announcements by the China Foreign Exchange Trading Center (CEFTC) to guide the onshore rate and curbed bilateral volatility by imposing trading band limits for the onshore RMB-USD[10]. Finally, in response to the financial crisis, the PBoC once again pegged its exchange rate against the USD.

Equation (1.2) is the supply equation. In particular, this equation describes the net supply of USDs arising from the consolidated dollar position across FX banks. Even though domestic agents and FX banks are individually constrained by regulation in their ability to hold dollars, in aggregate, this supply function must depend on the exchange rate. Further, for a constant exchange rate the supply of USDs in the onshore FX interbank market also depends on China's trade position since this is one of the major sources of USDs inflows. Finally, equation (1.3) describes the dependence of the trade balance on the other endogenous variables. Cleary, it will depend on both the exchange rate since the two are jointly determined in equilibrium and FX interventions as suggested by Jeanne (2012) and Benigno and Fornaro (2012).

1.2.2 Exchange rate regimes in China

In figure 1.1, I have plotted the time-series of the onshore RMB-USD rate for the sample period under study. From visual inspection it is clear that the most prominent feature of

[9]I implicitly assume in the VAR specification that this policy path is accurately captured by the lagged realized values of the exchange rate or other lagged variables. Regressing the end of month central parity rate on the lagged end of month realized onshore rate the R^2 is of 99.6%

[10]For example, see figure 2 in Frenkel and Wei (2007). The bilateral exchange rate volatility against the dollar is far less pronounced than that against the Euro and the Yen.

the data is the shift in the exchange rate regime occurring on July 21st 2005. On this date, the PBoC revalued the Chinese currency against the USD by approximately 2% after having held it fixed at 8.28 RMB/USD since 1997. Further, the PBoC also publicly announced that it had shifted its exchange rate from a *de facto* peg to the USD to a managed float. Thus, the period under study presents a clear structural break in policy occurring around the middle of the sample that divides it into two distinct exchange rate regimes. The first regime extends from January 1998, the start of my sample, to June 2005 and throughout the exchange rate remains fixed. From July 2005 onward, instead, the PBoC has implemented a managed float around a gradual and highly predictable appreciation path[11].

Though the July 2005 revaluation is the most prominent feature of the sample, two other policy interventions occurred in the period under study. First, a caveat to my dichotomous characterization of the sample is represented by the period between October 2008 and May 2010. During this period, the PBoC effectively re-pegged the RMB to the dollar in response to the global financial crisis and the steep fall in world trade. Further, the spot rate volatility declined sharply. Second, in conjunction with the July 2005 revaluation, the PBoC established daily fluctuation limits ($\pm 0.3\%$) for the onshore RMB-USD rate around a central parity rate announced each day by the CEFTC. In the managed float sample I study, the PBoC widened these bands twice. A first change was implemented in May 2007 widening the trading band to $\pm 0.5\%$. A second in April 2012 widening it to $\pm 1\%$[12].

This change in the daily trading band limits may affect the monthly volatility in the RMB-USD rate making it unstable over time. In practice, I find that at the monthly frequency the point estimates of the residual variances in these three samples are almost identical[13]. This is likely due to the fact that the trading bands govern intraday variation of the RMB-USD rate. The CEFTC, instead, controls day to day changes of the central par-

[11]Note that this distinction in the exchange rate regime is *de facto*. *De jure*, the regime has been a managed float throughout the period under analysis. For details, a PBoC statement on the policy stance is available here

[12]A third adjustment was made in March 2014 which is however outside of my sample period

[13]The estimated conditional standard deviations are respectively 0.46%, 0.45% and 0.41%.

ity rate and through this mechanism the month to month RMB-USD rate variation. It is, in fact, the month to month variation in these announced rates that explains the vast majority of the onshore RMB-USD movements[14] at the monthly level rather than the intraday variation that is susceptible to the trading band limits.

In this paper, therefore, I exploit the July 2005 exchange rate regime shift to identify my econometric model. In particular, I assume that the change in the reduced form moments across regimes stems only from a break in the parameters governing the policy equation of the PBoC. I ignore changes in the daily trading band limits as these don't seem to impact my model which is specified at the monthly level. With regards to the crisis period, I initially include this period in the managed float regime. However, in section 1.6.1 I perform various robustness checks to show how this assumption influences my results.

1.3 Identification

In this section, I first illustrate the problem of identification in structural VARs. Second, I show that under my identifying assumptions, the policy shock of interest is identified. Third, I discuss the plausibility of my identifying assumptions. Fourth, I highlight similarities and differences between my identification strategy and those adopted by the existing literature.

1.3.1 The VAR Identification Problem

Consider a general p lag structural VAR form:

$$A_0 Y_t = \kappa + A_1 Y_{t-1} + ... + A_p Y_{t-p} + \epsilon_t \tag{1.4}$$

where Y_t is $N \times 1$ data vector, κ is $N \times 1$ vector of constants, A_i is a $N \times N$ matrix of structural coefficients and the structural shocks, ϵ_t are distributed $N(0, D)$, where D is diagonal. The

[14] At the monthly level, the variation in the central parity rate accounts for almost 90% of the variation in the onshore RMB-USD exchange rate. Further, the difference between the two standard deviations is in the order of .05%.

reduced form of (1.4) may be expressed as:

$$Y_t = A_0^{-1}\kappa + A_0^{-1}A_1Y_{t-1} + ... + A_0^{-1}A_pY_{t-p} + A_0^{-1}\epsilon_t \tag{1.5}$$

$$Y_t = c + B_1Y_{t-1} + ... + B_pY_{t-p} + u_t \tag{1.6}$$

where c is vector of constants, B_i is a $N \times N$ matrix of reduced form coefficients and the reduced form errors, u_t, are distributed $N(0, \Sigma)$.

The VAR identification problem is the impossibility to recover uniquely the matrices governing (1.4) from (1.6). This lack of identification can be illustrated by examining the reduced form covariance term Σ. Consider a matrix pair (QA_0, QDQ^T) such that $QA_0 \neq A_0$, $QDQ^T \neq D$ and Q is a full rank matrix. By definition:

$$\Sigma = (A_0^{-1}Q^{-1})(QDQ^T)(A_0^{-1}Q^{-1})^T \tag{1.7}$$

$$= A_0^{-1}Q^{-1}QD(A_0^{-1}Q^{-1}Q)^T \tag{1.8}$$

$$= A_0^{-1}D(A_0^{-1})^T \tag{1.9}$$

so that premultiplication of (1.4) by a full rank matrix Q will yield the same reduced form as the original model (A_0,D). More precisely, there exists an uncountable set of matrix pairs ($\tilde{A}_0, \tilde{D}$), each corresponding to a different structural VAR, that satisfy (1.7). Thus, without further assumptions, (A_0, D) cannot be recovered from the reduced form covariance term alone.

An equivalent implication of equations (1.5) and (1.6) is that the reduced form errors may be expressed as a linear combination of the structural shocks (and vice versa):

$$u_t = A_0^{-1}\epsilon_t \tag{1.10}$$

Knowledge of A_0 identifies the structural shocks in (1.10) and, as above, the structural form in (1.4). In this case, the VAR will be fully identified since all structural shocks are recovered from the reduced form errors. However, (1.10) also implies that knowledge of the i-th row of A_0 alone recovers the i-th structural shock. In this case, (1.4) is partially identified since the i-th structural shock is recovered, while the model is agnostic about the remaining $N-1$ shocks.

1.3.2 Identification of the policy shocks

In this section I establish that the policy shock is identified. I do this in two steps. First, under my identifying assumptions a submatrix of the reduced form covariance matrix is a function of strict subset of the parameters governing the structural VAR. Crucially, this subset includes the parameters of the policy equation. Second, there exists a mapping from the reduced form moments in the two submatrices - one per regime - to this subset. This mapping only identifies the parameters in the policy equation, and therefore, my structural VAR is partially identified.

The structural VAR form of my model is:

$$\underbrace{\begin{bmatrix} 1 & -\alpha_1^{(i)} & -\alpha_2^{(i)} \\ -\beta_1^{(i)} & 1 & -\beta_2 \\ -\gamma_1^{(i)} & -\gamma_2 & 1 \end{bmatrix}}_{A_0^{(i)}} \underbrace{\begin{bmatrix} s_t \\ I_t \\ TB_t \end{bmatrix}}_{Y_t} = \kappa^{(i)} + A_1^{(i)} Y_{t-1} + ... + A_p^{(i)} Y_{t-p} + \underbrace{\begin{bmatrix} \epsilon_t^D \\ \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}}_{\epsilon_t} \tag{1.11}$$

where the shocks ϵ_t are distributed $N(0, D^{(i)})$, and $D^{(i)}$ is a block diagonal matrix[15]. In fact, I assume that the policy shock ϵ_t^D is fundamental in nature and therefore orthogonal to both $\epsilon_{1,t}$ and $\epsilon_{2,t}$. However, I allow $\epsilon_{1,t}$ and $\epsilon_{2,t}$ to be correlated to each other. In (1.11), the superscript (i) indexes the parameters which are allowed to vary across exchange rate regimes. In particular, the parameters governing the policy reaction - $\alpha_1^{(i)}, \alpha_2^{(i)}, \sigma_D^{(i)}$ - are allowed to shift across regimes. Further, the parameters $\beta_1^{(i)}, \gamma_1^{(i)}$ as well as all intercepts, $\kappa^{(i)}$, and lagged effects, $A_j^{(i)}$, may be different before and after the policy shift.

[15]The first upper left block is a scalar containing the variance of the policy shock, $\sigma_D^{2,(i)}$. The second block is 2×2 matrix with the variances σ_1^2 and σ_2^2 on the diagonal and the covariance term $\sigma_{1,2}$ off diagonal.

For notational purpose, define $\Phi^{(i)}$ to be the set of parameters that govern the matrices $(A_0^{(i)}, D^{(i)})$. Define the scalar valued functions $g(\bullet)$ and $h^{(i)}(\bullet)$ to be:

$$g(\bullet) = \frac{\sigma_1^2 + \beta_2^2\sigma_2^2 + 2\beta_2\sigma_{1,2}}{(1-\beta_2\gamma_2)^2} \tag{1.12}$$

$$h^{(i)}(\bullet) = \frac{\beta_1^{(i)} + \beta_2\gamma_1^{(i)}}{1-\beta_2\gamma_2} \tag{1.13}$$

note that $g(\bullet)$ is a function of parameters which are assumed to remain constant across regimes. On the other hand, $h^{(i)}(\bullet)$ depends on parameters that may vary across regimes. Finally, let $\Sigma^{(i)}$ denote the reduced form covariance matrix of (1.11). Denote with $\Sigma_S^{(i)}$ the 2×2 upper left submatrix of $\Sigma^{(i)}$ that includes the variance of the exchange rate, the variance of FX intervention and their covariance.

Claim: Given the model in (1.11) and the identifying assumptions listed below, it possible to identify the effect of a policy shock on the exchange rate, FX intervention and the trade balance in the managed float regime.

Assumption 1: $\alpha_2^{(i)} = 0$ in both regimes. I assume that the PBoC contemporaneously only reacts to deviations of the exchange rate from its target.

Assumption 2: $\alpha_1^{(1)} = 0$ in the fixed exchange rate regime. I assume that the PBoC had a perfectly elastic demand for FX at the target exchange rate.

Assumption 3: $\sigma_D^{(1)} = 0$ in the fixed exchange rate regime. There were no policy shocks in the implementation of the fixed exchange rate target.

Assumption 4: $g(\bullet)$ is constant across regimes.

Assumption 5: $\alpha_1^{(2)} \leq 0$ in the managed float regime. The PBoC accommodates shocks by leaning against the wind.

From (1.10), identification of the policy shock requires knowledge of the policy parameters $\alpha_1^{(i)}, \alpha_2^{(i)}, \sigma_D^{(i)}$. However, by assumption $\alpha_2^{(1)}$, $\alpha_2^{(2)}$, $\alpha_1^{(1)}$ and $\sigma_D^{(1)}$ are all known. Therefore the only policy parameters left to recover are $\alpha_1^{(2)}$ and $\sigma_D^{(2)}$. I show below how these can be recovered.

Sketch of Proof: In general, the elements of $\Sigma^{(i)}$ will be a function of all elements in $\Phi^{(i)}$. The key to identifying the remaining unknown policy parameters $\alpha_1^{(2)}, \sigma_D^{(2)}$ is to realize that, under assumption 1, it possible to express the submatrix $\Sigma_S^{(i)}$ as a function of only the policy parameters $\alpha_1^{(i)}, \sigma_D^{(i)}$ and the functions $h^{(i)}(\bullet)$ and $g(\bullet)$. Thus, $\Sigma_S^{(i)}$ may be expressed as[16]:

$$\Sigma_S^{(i)} = \begin{bmatrix} \Sigma_{1,1}(\alpha_1^{(i)}, \sigma_D^{(i)}, h^{(i)}(\bullet), g(\bullet)) & \Sigma_{1,2}(\alpha_1^{(i)}, \sigma_D^{(i)}, h^{(i)}(\bullet), g(\bullet)) \\ \Sigma_{2,1}(\alpha_1^{(i)}, \sigma_D^{(i)}, h^{(i)}(\bullet), g(\bullet)) & \Sigma_{2,2}(\alpha_1^{(i)}, \sigma_D^{(i)}, h^{(i)}(\bullet), g(\bullet)) \end{bmatrix} \tag{1.14}$$

In light of the exchange rate regime shift, the model in (1.11) implies that reduced form covariance matrix may differ across regimes. Let $\Sigma^{(1)}$ denote the covariance matrix in the fixed exchange period and $\Sigma^{(2)}$ that in the managed float period. Both of these matrices will contain submatrices $\Sigma_S^{(i)}$ like (1.14), since assumption 1 holds in both periods. Further, under assumptions 1,2 and 3, the variance of the exchange rate and the covariance with FX intervention are both equal to zero[17] in the fixed exchange rate period. Thus the submatrix $\Sigma_S^{(1)}$ may be expressed as:

$$\Sigma_S^{(1)} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{2,2}(g(\bullet)) \end{bmatrix} \tag{1.15}$$

This occurs because I have assumed that (i) the PBoC has perfectly elastic demand for FX at the target exchange rate ($\alpha_1^{(1)} = 0$), and (ii) there are no policy surprises in the fixed exchange rate regime ($\sigma_D^{(1)} = 0$). In this policy regime, therefore, there is no variation in the exchange rate since (i) implies that the exchange rate is exogenous and (ii) that the

[16]Derivations are given in Appendix
[17]Derivations are given in Appendix.

exogenous variation is zero. Further, FX intervention is entirely passive since the PBoC only makes FX purchases to implement the peg. As a result, the residual variation in FX intervention directly identifies $g(\bullet)$. Intuitively, the function $g(\bullet)$ measures the total variation in FX intervention that arises from shocks other than the policy shock. Since there are no policy shocks in the fixed exchange rate regime and policy is entirely passive, it must be that $g(\bullet)$ is identified by the reduced form of (1.15). Further, because there is no variation in the exchange rate, the parameters $\beta_1^{(1)}, \gamma_1^{(1)}$ cannot be identified. As a consequence, nothing may be learned about $h^{(1)}(\bullet)$ from (1.15) either and the parameter is unidentified.

The parameters $\alpha_1^{(2)}, \sigma_D^{(2)}$ can be recovered by noting that there exists a mapping between the four reduced form moments in $\Sigma_S^{(1)}$ and $\Sigma_S^{(2)}$ and the four unknown parameters $\alpha_1^{(2)}, \sigma_D^{(2)}, h^{(2)}(\bullet), g(\bullet)$. Solving this system of non-linear equations produces estimates of the policy parameters $\alpha_1^{(2)}, \sigma_D^{(2)}$, as well as values for $h^{(2)}(\bullet)$ and $g(\bullet)$. From above, it follows that knowledge of $\alpha_1^{(2)}, \sigma_D^{(2)}$ in conjunction with assumption 1 identifies the policy shock in the managed float period. On the other hand, knowledge of $h^{(2)}(\bullet)$ and $g(\bullet)$ alone is insufficient to identify the underlying parameters of these functions, so the other shocks remain unidentified. Thus the VAR is partially identified.

This non-linear equation system has two properties: (i) its solution is generally not unique due to its quadratic nature, (ii) a solution does not exist for all matrix pairs $(\Sigma_S^{(1)}, \Sigma_S^{(2)})$. Assumptions 3 is sufficient to guarantee uniqueness of the solution in a non-trivial subset of the parameter space. Independently of assumption 3, when $\Sigma_S^{(1)}$ and $\Sigma_S^{(2)}$ are estimated independently it is possible that no solution exist to the equation system. This occurs because the estimate of $g(\bullet)$ implied by $\Sigma_S^{(1)}$ lies outside the feasible space of parameters given $\Sigma_S^{(2)}$. Neither of these issues bind for the baseline results in section 1.6, however, the lack of uniqueness may bind when using a very diffuse prior as I do in section 1.6.1.

Finally note that despite knowledge of the policy parameters in the fixed exchange rate regime, the effect of a policy shock is not identified. In fact, in the fixed exchange rate regime, the policy variance is zero. Thus, without independent variation in FX interventions, the IRFs of the endogenous variables to the policy shock remain unidentified. A count of the full set of moments and unknowns suggests that the entire model, in fact, may be identified. I choose not pursue this for the following reason. Intuitively, policy shocks are identified because policy is entirely passive in the fixed exchange rate regime and it is not in managed float. This distinction identifies the policy shock from the other two shocks but not the other two shocks from each other. In a small VAR, such as the one I estimate, it is thus preferable to impose as little structure on these remaining shocks as possible. In fact, it may be that the remaining reduced form shocks cannot be orthogonalized into a set of structural shocks each with a clear, distinct interpretation[18].

1.3.3 Identifying assumptions

I now proceed to discuss the plausibility of the identifying assumptions.

Assumption 1: $\alpha_2^{(i)} = 0$ in both regimes. I assume that the PBoC contemporaneously only reacts to deviations of the exchange rate from its target.

In the fixed exchange rate regime, the assumption holds since the exchange rate is exogenously determined and fixed. In the managed float regime assumption 1 is sensible for the following reasons. First, to the extent that the PBoC's objective is to stabilize the exchange rate (Chang et al. (2013)), FX interventions should only respond to deviations from the PBoC's target rate. Any attempt to use FX purchases to stabilize both the exchange rate and the trade balance inherently trades off volatility in the exchange rate against volatility in the trade balance. Second, while the stabilizing effect of FX purchases on the exchange rate is immediate, a nominal depreciation may, at least initially, deteriorate the trade balance before inducing an improvement. Thus, at least in the short-run, it is unclear to what extent FX intervention is an effective instrument to influence the trade balance. Third, in an environment with information lags the PBoC may not directly react to the trade balance simply because it does not contemporaneously observe it (or it observes it imprecisely).

[18]For example, one may think of the shocks $\epsilon_t^1, \epsilon_t^2$ as combinations of fundamental shocks. If these combinations overlap then ϵ_t^1 and ϵ_t^2 will be in fact correlated.

These three reasons suggest that, within the month, the PBoC uses FX interventions only (or predominantly) to stabilize the exchange rate. However, it chooses the policy path of the exchange rate in function of lagged realizations of the trade balance. For example, the PBoC chooses to steepen the appreciation path in response to a protracted period of large trade surpluses or to slow it in response to a series of negative shocks.

Assumption 2: $\alpha_1^{(1)} = 0$ in the fixed exchange rate regime. I assume that the PBoC had a perfectly elastic demand for FX at the target exchange rate.

Assumption 3: $\sigma_D^{(1)} = 0$ in the fixed exchange rate regime. There were no policy shocks in the implementation of the fixed exchange rate target.

Assumptions 2 and 3 are made to reflect the state of policy in the fixed exchange rate regime. By definition, under a fixed exchange rate regime, any supply or demand imbalance in the onshore FX interbank market must be fully accommodated. Thus demand must be perfectly elastic at the target rate. Further, exogenous FX intervention does not occur because it is incompatible with the fixed exchange rate target of the PBoC. In fact, *ceteris paribus*, exogenous FX intervention by the PBoC would have driven the onshore rate away from the target exchange rate. Note that in the data, the RMB-USD exchange rate is not entirely fixed. The residual standard deviation of the RMB-USD rate between January 1998 and July 2005 is of 0.05137%. If I exclude observations prior to January 1999, this conditional standard deviation falls to 0.008407%. The difference between the two estimates is driven by three observations in the first year of the sample in which the bilateral exchange rate moved from 8.2770 to 8.28 RMB/USD (or vice versa). In both cases, the residual variance is sufficiently small that my assumptions are, to a first approximation, reasonable.

Assumption 4: $g(\bullet)$ is constant across regimes

This the potentially controversial element of my identification strategy. This assumption is a crucial step in my identification strategy as it links the reduced forms across regimes. Without this assumption, the simultaneity between the spot rate and FX interventions in the managed float period cannot be disentangled since the policy shift contains no identifying information. The fundamental concern with this assumption is that the shift in exchange rate regime may also alter the behavior of economic agents (Lucas (1972)). I argue below, however, that the size of the policy shift, capital controls and the PBoC's policy in the aftermath of the revaluation in great part attenuate these concerns. As well, I investigate the extent to which violations of assumption 4 affect the results. I assume that the value of $g(\bullet)$ across regimes is related by some random variable ς. The choice of prior for this random variable can be used to model the researcher's uncertainty around assumption 4. For example, an infinitely diffuse prior is equivalent to rejecting it, while a relatively tight prior centered around 1 assumes that the assumption almost holds. The results of this exercise are discussed in section 1.6.1.

The vast majority of USD inflows into the onshore FX interbank market are tied to capital inflows and trade inflows into China. Thus, my identification strategy fails to the extent that the policy shift affects the conditional volatility of these net inflows. However, in standard open economy models (Vegh (2013)), the exchange rate regime and fundamental volatility are usually separately modeled. By construction, in these models, a change in the exchange rate regime affects only the intensity with which FX interventions and/or the exchange rate adjust in response to fundamental shocks. Shock volatility is unchanged across regimes since it is exogenous. As well there is empirical evidence that the volatility of macro aggregates does not change when moving from fixed to floating (Duarte (2003)). In my application, therefore, it seems equally reasonable to assume that, for example, real-side shocks, such as productivity or demand shocks, affecting the onshore USD supply may be well described by such a stylized model. However, a concern still remains that the policy shift may have affected market participant's incentive to increases speculative pressures on the onshore RMB-USD rate. In particular, their incentive to invest in onshore RMB denominated assets.

Before and after the July revaluation, the RMB remained substantially undervalued against the USD according to market participants. Thus incentives to hold long-term RMB positions were largely unchanged in response to the policy shift. However, the change in the exchange rate regime may have intensified speculative USD inflows by investors who perceived the probability of another imminent revaluation to be higher after July 21, 2005. The extent to which this increase in speculative activity may affect the net USD onshore supply volatility, however, is limited by two factors: first, existing capital account regulations in China limit the ability of foreigners to both obtain RMB assets in exchange for FX and to quickly reverse cross-border positions. For example, assets purchased under the Qualified Foreign Institutional Investor (QFII) program - the main investment vehicle for foreign financial entities in China - may not be repatriated prior to a predetermined amount of time[19]. Further, each authorized financial institution is subject to an investment quota limit. Though these regulations may be circumvented, their mere existence certainly hinders speculative activity as they have allowed the PBoC to keep control of the nominal exchange rate, despite a certain amount of monetary policy independence. Second, the incentive for foreign investors to circumvent the capital controls is directly related to the probability that another large revaluation occur in the period subsequent to the July 21st, 2005 intervention. However, in the days successive to the shift in the exchange rate regime the PBoC clearly signaled it would not allow large and sharp changes in the bilateral RMB-USD rate. In the PBoC release of July 21, 2005 itself, the PBoC emphasized its responsibility to keep the "[...] the RMB exchange rate *basically stable* at an adaptive and equilibrium level [...]"[20]. As well, the Chinese official media attempted to cool appreciation expectation by noting that " [...] expectation for a bigger appreciation of the yuan's vale was, and will be, unrealistic. [...]" as reported by the China Daily editorial of July 22, 2005. Finally, currency traders reported that the PBoC intervened heavily on July 22, 2005 to keep the currency from appreciating and hitting the top of its trading band. Such an event, in the mind of many market participants, would have triggered expectations of further appreciation of the currency. To the extent that the PBoC was successful in coordinating expectations, this

[19] Lock up periods typically range from 3 months to 1 year counted from the day the principal is remitted in full. Further, QFII status has to be granted by the State Administration for Foreign Exchange (SAFE) so that at any point in time the number of foreign investors is fixed.

[20] The quote is from the Public Announcement of the People's Bank of China on Reforming the RMB Exchange Rate Regime dated July 21st, 2005 available here. Emphasis is added.

mechanism also acts to constraint the magnitude of the change in the USD inflow volatility.

Further, while capital controls and expectation coordination act to limit the change in the USD inflow volatility for a given size of the policy shift, I also argue that the policy shift in question was itself small. In the spectrum of exchange rate regimes (Levy-Yeyati and Sturzenegger (2005)), the policy change can be characterized as a rather minor move towards a more flexible arrangement. In fact, the conditional monthly standard deviation of the RMB-USD rate in the managed float regime remained low[21] and paired with substantial FX intervention. A small change in the policy stance is also consistent with the history of economic reforms of the country and policy statements of Chinese officials. In fact, it is conventional wisdom that Chinese authorities prize stability and implement gradual policy change (Frenkel and Wei (2007)). To the extent that a small regime discontinuity also induces small expectation-formation effects on the model parameters as in Leeper and Zha (2003), then inference in my model remains approximately correct.

Assumption 5: $\alpha_1^{(2)} \leq 0$ in the managed float regime. The PBoC accommodates shocks by leaning against the wind.

Assumption 5, in my view, reflects the consensus of the FX intervention literature. In fact, empirically, researchers tend to find that Central Banks across countries lean against the wind (Sarno and Taylor (2001)). The assumption is highly plausible also from a theoretical perspective. In fact, if α_1 were strictly greater than zero this would imply that the PBoC uses FX interventions to accentuate the effect of shocks on the exchange rate. Given the explicit goal of the PBoC to maintain a stable and competitive exchange rate (Chang et al. (2013)) this seems unreasonable. Note that assumption 5 does not imply that the impact effect of FX intervention on the exchange rate is positive. In a simple bivariate context, assumption 5 is a restriction on the slope of the demand curve. The effect of exogenous FX intervention depends on the slope of the supply curve.

[21] I estimate that the conditional monthly standard deviation of the RMB-USD rate was of only 0.4% in the period subsequent to July 2005. The corresponding amount for free floating currency pairs such as the Euro-USD and the USD-GBP is of 2.31% and 2.05% respectively.

1.3.4 Relationship to the literature

The identification strategy employed in this paper is related to two lines of work: (i) the empirical literature on FX interventions, in particular to the work of Kearns and Rigobon (2005), (ii) the literature on VARs identified by structural change (Rigobon (2003), Bacchiocchi and Fanelli (2012), Bacchiocchi et al. (2014)). In all three instances, structural change either in the form of a change in intervention probability (Kearns and Rigobon (2005)), conditional volatility (Rigobon (2003)) or both (Bacchiocchi and Fanelli (2012)) is exploited to identify the econometric model in question.

In these models, a break in a subset of the structural parameters - e.g. the policy function parameters - is assumed to rationalize the observed change in the reduced form. My identification procedure shares this characteristic but differs in that it makes additional assumptions about the structure of the model. In particular, these assumptions are sufficient to directly identify a subset of the structural parameters of the model unaffected by the structural change. Knowledge of these structural parameters is then employed to disentangle the endogenous relationship(s) that characterizes the second policy regime. In this paper, assumptions 1-3 guarantee that the residual variance of FX interventions in the fixed exchange rate regime resolves the simultaneity in the managed float regime. These assumptions, as well as assumption 5, are in addition to the assumption that only a subset of parameters changes across regimes.

My approach is more restrictive than that of Kearns and Rigobon (2005) and Bacchiocchi and Fanelli (2012) since it requires additional identifying assumptions. It has, however, four distinct advantages *vis à vis* their approaches in this application. First, my identifying assumptions may be sufficient, and are in this application, to guarantee global uniqueness of the solution to the non-linear equation system. Bacchiocchi and Fanelli (2012)'s rank condition (Proposition 1, pg. 14) instead is necessary and sufficient only for local identification. Multiplicity of solutions can, in practice, have a serious impact on inference (e.g. multiple peaks in the posterior likelihood). Second, my approach only identifies the effect of the policy shock, it remains agnostic about the other shocks in the VAR. This is not the case in Rigobon (2003) and Bacchiocchi and Fanelli (2012) which identify the entire VAR system. Third, with the macroeconomic time-series data available to us, identification methods that rely on structural change typically require large changes in the reduced forms for the model to be well identified in practice. The policy shift of July 2005 may not satisfy this condition and my identification strategy, in this instance, may have more traction on the problem since it effectively relies on a subset of the reduced form estimates to achieve identification. Obviously, this sharper identification comes at the expense of additional assumptions which may or may not be appropriate. Fourth, the simplicity of my approach is attractive because it gives the identification strategy an intuitive interpretation.

1.4 Estimation and Inference

The reduced form of (1.11) is an unrestricted 3 variable VAR in the log onshore RMB-USD rate, FX intervention and the trade balance. This may be expressed as:

$$Y_t^{(i)} = B_0^{(i)} + B_1^{(i)} Y_{t-1}^{(i)} + ... + B_p^{(i)} Y_{t-p}^{(i)} + \nu_t^{(i)} \tag{1.16}$$

where (i) indexes the exchange rate regime and $B_j^{(i)}$ are the coefficient matrices for each regime. In the fixed exchange rate period, the first variable in $Y_t^{(1)}$, the log of the spot rate, empirically displays only infinitesimally small variation. Thus only the reduced form equations involving FX intervention and the trade balance are estimated. The coefficients and residual variances in these equations are consistently estimated because the omitted regressor - the USD-RMB rate - has zero variation. In the managed float regime, all equations in (1.16) are estimated.

I estimate independently both systems of equations using a Bayesian procedure. There are two source of joint inference I ignore by doing this. The first is that the two samples in part overlap so that observations from the fixed exchange rate regime are the starting conditions for the managed float regime. I address this by cutting the samples so that there is no overlap in the data. Second, when estimating the covariance matrices independently it is possible that the model be overidentified. For the model to be true with certainty, joint inference must be performed. However, even when the covariance matrices are estimated independently I find that the model is overidentified in a only a trivial amount of draws. I therefore estimate the two VAR systems separately and discard any draw for which the model is overidentified. Finally note that given the structure in (1.11) and assumption 1-5, there is no common inference on the $B_j^{(i)}$ reduced form parameters that can be made across regimes.

In both regimes, the system in (1.16) may be equivalently expressed in its simultaneous equations form:

$$Y^{(i)} = Z^{(i)}\Gamma^{(i)} + E^{(i)} \tag{1.17}$$

I estimate the VARs using a Normal - Inverse Wishart prior:

$$vec(\Gamma^{(i)})|\Sigma^{(i)} \sim N(vec(\underline{\Gamma}^{(i)})|\Sigma^{(i)} \otimes \underline{N}^{-1,(i)}) \tag{1.18}$$

$$\Sigma^{(i)} \sim IW_n(\underline{v}^{(i)}\underline{S}^{(i)}, \underline{v}^{(i)}) \tag{1.19}$$

where n indexes the dimension of the reduced form VAR and $(\underline{\Gamma}^{(i)}, \underline{S}^{(i)}, \underline{N}^{-1,(i)}, \underline{v}^{(i)})$ are hyperparameters governing the prior. These are chosen so that the prior be uninformative (Uhlig (2005)). The posterior distribution and sampling of the reduced form parameters $(\Gamma^{(i)}, \Sigma^{(i)})$ are straightforward due to the conjugate nature of the Normal-Inverse Wishart prior.

Let $\sigma_{\nu_2}^{(1)}$ denote the standard deviation of the reduced form residual in the FX intervention equation from the fixed exchange rate regime. To sample this quantity, sample $\Sigma^{(1)}$ from the IW posterior of period 1 and extract the appropriate variance term from the covariance matrix. To identify the model I use the following procedure:

1. I take a draw from the posterior distribution of $\sigma_{\nu_2}^{(1)}$

2. I take a draw from the posterior distribution of $(\Gamma^{(2)}, \Sigma^{(2)})$

3. For each draw pair, I solve for the parameters as described in section 1.3.2

4. Save the resulting IRF draw ($\widetilde{\Theta}$) and structural model probability $p(\widetilde{\Theta}|Y^{(2)})$

Repeating these steps simulates the posterior distribution of the structural parameters. As shown by Inoue and Kilian (2013), point-wise confidence intervals tend to misrepresent the shape of the IRFs and have poor nominal coverage. Therefore, I report instead the posterior mode and the associated credible set[22] in my empirical analysis. The posterior likelihood function of the structural parameters of a SVAR model may be written as:

$$p(\Theta|Y^{(2)}) = |J|p(\Gamma^{(2)}, \Sigma^{(2)}|Y^{(2)})p(Q|\Gamma^{(2)}, \Sigma^{(2)}) \tag{1.20}$$

$$= |J|p(\Gamma^{(2)}, \Sigma^{(2)}|Y^{(2)})p(\sigma^{(1)}_{\nu_2}|\Sigma^{(2)}, Y^{(1)}) \tag{1.21}$$

where $|J|$ is the appropriate Jacobian matrix. In general, the prior over the rotation matrix Q is chosen to be flat and independent of the reduced form parameters. However, under my identifying assumptions, knowledge of $\sigma^{(1)}_{\nu_2}$ given $\Sigma^{(2)}$ is sufficient to pin down the rotation matrix Q. Thus the probability of any Q rotation, in fact, depends on the prior over $\sigma^{(1)}_{\nu_2}$ for a given draw of the period 2 reduced form. Under assumptions 1-5, this prior is the posterior of the fixed exchange rate period FX intervention conditional variance.

The full data sample covers the period between 1998:M1 to 2013:M12, covering exactly 16 years of data or equivalently 192 months. Period 1, the fixed exchange rate regime, includes 1998:M1 to 2005:M6 for a total of 90 monthly observations. The first reduced form bivariate VAR in FX interventions and the trade balance is estimated on data from this period. Period 2, the managed float regime, spans 2005:M7 to 2013:M12 for a total of 102 monthly observations. The second specification is estimated using this data. The start of the sample is limited by the fact that prior to 1998 balance of payments data is unavailable. The data for this paper is obtained from three main sources. The bilateral RMB-USD rate is obtained from the H.10 release of the Federal Reserve Board. Trade balance data

[22]For ease of interpretation, the credible set here is formed as the outer contour of the credible set.

is obtained from the Directions of Trade Statistics (DOTS) dataset of the IMF. Data on FX holdings and balance of payments is obtained from the PBoC via Haver Analytics.

A last issue regards how to introduce FX interventions and the trade balance in the econometric specification. This is quite a crucial issue, since the volatility of USD inflows is assumed to be constant across the sample. Clearly, FX interventions and the trade balance cannot be measured in either nominal or real USDs as the size the Chinese economy has radically changed over the sample period. Similarly, the size of shocks measured in USDs must have also changed. Scaling interventions by GDP is not appropriate either, since the openness of the Chinese economy has also changed substantially over time. I therefore normalize FX interventions and the trade balance by the average level of net transactions with the rest of the world. I measure this as the two-sided 12-month moving average of the Chinese current account. This deflator well captures the rise and subsequent fall, albeit to a higher level, in China's net transactions with the rest of the world.

1.5 Measuring FX intervention

One of the major challenges for researches studying FX intervention policy has been the reluctance of Central Banks to release detailed information about their activities. In particular, information on the currency composition and maturity structure of their FX holdings is rarely known. Though disclosure standards worldwide have been rapidly evolving (see Adler and Tovar (2011)) and for selected countries these data are now available, measuring policy induced changes in FX reserves remains challenging, especially so in emerging markets. This distinction between active and passive changes in reserves has been recently re-emphasized by Dominguez et al. (2012) and Dominguez (2012) in the context of the financial crisis.

For the purpose of this paper, I am interested in recovering active changes in FX assets. This component of reserves is the focus of this paper as it is what researchers typically associate with FX intervention. Given the lack of direct measures of policy, it has been common in the literature to proxy reserve policy by using differences in FX reserve stocks. However, changes in this measure can be expressed as:

$$\Delta FX_t = \underbrace{I_t}_{\text{FX intervention}} + \underbrace{\Delta^p FX_t}_{\text{Valuation effects}} + \underbrace{r_t FX_{t-1}}_{\text{Interest Accrual}} \tag{1.22}$$

so that changes in reserve stocks are a noisy measure of policy as they also include valuation effects and interest accrual. Flow data is a better measure of FX interventions as it excludes by construction the valuation effect. This valuation effect is, in practice, the real obstacle to measuring FX intervention precisely. However, discrepancies between flow and stock data may also infrequently arise due to transfers. In the case of China, the PBoC has, on occasion, employed FX assets to recapitalize large state-owned financial institutions. Prior to 2007, these recapitalizations occurred through the Central Huijin Investment Company, the predecessor of China Investment Corporation, with investment capital supplied by the PBoC. A list of these recapitalizations is provided in table 1.2. The list only includes recapitalizations that were undertaken using FX reserves held by the PBoC[23]. By adding back these transfers to the stock of FX holdings of the PBoC I correct for this source of discrepancy. In practice, the importance of this adjustment is very limited[24].

Dominguez et al. (2012) construct their measure of FX interventions by estimating the valuation and interest accrual components and netting them out from the change in FX stocks. Interest accrual is estimated using data contained in the SDDS dataset[25]. The SDDS dataset is useful to estimate interest accrual as it breaks out FX holdings into deposits and securities. It does not, however, have any information about the maturity of FX holdings and, more importantly, about their currency composition. They therefore use aggregate information available from the Currency Composition of Official Foreign Exchange Reserves (COFER) dataset to calibrate the currency portfolios. In this paper I cannot follow this approach for two reasons: (i) China does not participate in the SDDS initiative, (ii)

[23]For more details see Zhang (2010).

[24]With the exception of one data point there is no substantial difference in the estimated latent factor. The data point in question is that associated with the recapitalization undertaken in December 2003

[25]The SDDS is one of IMF's Data Dissemination Standards initiatives to which countries participate is on a voluntary basis

even for countries in the SDDS, data is only available from around the mid 2000s so that the resulting time-series are short.

The approach taken in this paper is to model FX intervention as a latent variable of which I observe two signals. This problem can therefore be cast into a state space form and I use the Kalman filter to produce an optimal online forecast of the latent state. The measure equations of this system may be written as:

$$\Delta FX_t = r_t FX_{t-1} + I_t + \Delta^p FX_t \tag{1.23}$$

$$FXBOP_t = r_t FX_{t-1} + I_t \tag{1.24}$$

where ΔFX_t is the change in the stock of FX reserves, $FXBOP_t$ is the flow of FX reserves observed in the balance of payments statistics and t indexes months. The first variable is observed on a monthly basis while the latter is observed at a quarterly frequency. Ignoring the interest accrual term, the filtering procedure I employ will attribute the total amount of FX intervention in a quarter measured by $FXBOP_t$ to the months within that quarter based on the noisy monthly information that is contained in ΔFX_t. This information is noisy because changes in ΔFX_t in any month may be due to valuation effects rather than FX intervention. Intuitively, the more precisely the valuation effect is approximated, the more precise will be the inference around the latent factor.

Since FX reserves are typically held in a small set of foreign currencies[26], valuation effects will depend on exchange rate fluctuations between reserve currencies and the USD. I therefore model valuation effects[27] as:

$$\Delta^p FX_t = \sum_{i \in \mathbf{C}} \dot{e}_t^i \beta_t^i FX_{t-1} + \nu_t \tag{1.25}$$

where $\mathbf{C}$ is the set of reserve currencies other than the dollar, FX_{t-1} is the stock of FX as-

[26]According to the COFER data the bulk of FX reserves in EMs is held in only 4 currencies: USD (64.9%), Euros (26%), GBP (4.4%) and Yen (2.1%).

[27]The exchange rate here is quoted as Dollars/LCU so that an increase is an appreciation of the LCU relative to the Dollar which implies that all else equal the dollar value of FX has increased.

sets and β_t^i is the value share of FX reserves in currency i[28]. Since the currency weights may vary over the sample, I employ information from the literature to calibrate priors for the value shares. In particular, Sheng (2013) finds evidence that except for the euro share all other currency weights do not fluctuate significantly over the period 1999-2007. I use this information to calibrate an informative prior for the β_t^i value shares. Details on the prior calibration and reserve currencies used to model the valuation effect are listed in table 1.1. In particular, I set all coefficients minus the euro share to be constant around Sheng's estimates. I model the break in the value share of the Euro to occur in September 2002. This is the month identified by Sheng to exhibit the portfolio shift. One caveat to my analysis is that Sheng's study only covers the period between 1999 to the end of 2007 so that if other portfolio shifts occur in the sample, my model will not be able to accommodate them.

In my empirical implementation, I define $(I_t + r_t FX_{t-1})$ to be the latent factor in my model rather than I_t directly. I do this because, differently than the valuation effect, my measures of FX intervention do not allow for clean separation of the interest accrual term from I_t even when aggregated at the quarterly level[29]. To identify I_t, I subtract my estimate of interest accrual from the latent factor. I proxy interest rates with the 10-year government bond yields as in Dominguez et al. (2012). Since the effective maturity of holdings is, in fact, unknown this might result in an imperfect approximation of the interest accrual term. In practice, accrued interest plays no significant role in determining the resulting time-series due to the scale of FX intervention in China.

The model has three sets of parameters: $|\mathbf{C}|$ currency shares, the variance parameters and the state equation autoregressive parameter(s)[30]. The currency weight priors are chosen to be informative as described above. The priors for the variance and AR parameters are chosen to be uninformative. All priors are chosen to be conjugate in nature. The model is estimated using standard Markov Chain Monte Carlo (MCMC) methods described by

[28]For a similar approach to this problem see Sheng (2013).

[29]The valuation effect instead is identified at the quarterly frequency as, by definition, it is the difference between $FXBOP_t$ and $\sum_{s=t-2}^{t} \Delta FX_s$ in the months t in which $FXBOP_t$ is observed.

[30]In this instance, the loading parameters are known and thus the variance parameters are separately identified. In general, this is not the case in a dynamic factor model.

Carter and Kohn (1994). I estimate the model using 5,000 burn in draws. I then save the following 10,000 draws to compute posterior means and standard deviations of the latent factor and parameters. In figure 1.2, I have plotted the estimated FX intervention time-series in blue and the associated standard errors in red. In green I have also plotted the changes in FX stocks. In parallel to the expansion of Chinese trade, FX intervention steadily rises from the start of the sample till the start of financial crisis. It remains lower in the subsequent period and recovers to levels similar to those of 2006/2007 by the end of the sample. The posterior estimates of the currency shares are also reported in table 1.1. These suggest that about 64.9% of Chinese FX assets are dollar denominated. The figure on the USD share is roughly in line with other estimates in the literature (55% - 65%, Prasad and Wei (2005)).

1.6 Empirical Results

Prior to estimation of (1.16), I use an information criterion to determine the number of lags to include. I do this separately for each exchange rate regime. I choose to specify both models with 3 lags. This is the maximum number of lags selected by any of the three information criteria used (AIC, HQIC, BIC) across the two regimes. I also do not choose to model (1.16) with a lower lag length because, by construction, the FX intervention variable features a form of smoothing of lag length 3. In fact, when I estimate the FX intervention variable, I assume that FX interventions measured from the stock data align with those measured from the balance of payments flow data. This effectively induces the Kalman Filter to smooth the within quarter variables so that they are consistent with the quarterly frequency observations.

I approximate the posterior distribution of the structural parameters by taking 500 draws of $\Sigma^{(1)}$, independently of $(\Gamma^{(2)}, \Sigma^{(2)})$, and 2,000 draws of $(\Gamma^{(2)}, \Sigma^{(2)})$. The posterior distribution of the parameters is therefore approximated by about 1,000,000 draws. The IRFs I have reported are limited, by choice, to the first 24 months after the impact period. With only 3 lags in the VAR specification and 102 monthly observations to estimate it, it is unlikely that inference from my model is reliable beyond the 1 to 2 year horizon. Given the sample size, including more lags is unlikely to solve this problem.

In figure 1.3, left column, I have plotted the estimated IRF of the onshore spot rate, FX interventions and the trade balance to an exogenous purchase of FX by the PBoC. In the right column of figure 1.3, I have plotted the corresponding variance decompositions. In blue I have plotted the modal response, and in two shades of red, respectively, the 68% and 90% confidence intervals. The size of the policy shock is normalized to induce a 1% depreciation in the onshore RMB-USD rate on impact. Thus the impact coefficient in the FX intervention IRF - .5941 - quantifies the size of the FX purchase made by the PBoC. Since, FX interventions in the model are measured as the ratio between FX interventions and the 12-month moving average of the current account, the size of such impact can only be measured at the mean of the scaling factor. I find that to induce a 1% depreciation of the spot rate, the PBoC must make purchases of 24.2 Bln USD and this purchase is statistically very different from zero. In fact, more than 99% of the posterior distribution for the impact effect lies above the zero. A purchase of 24.2 Bln USD within a month may seem extremely large but this in part depends on the fact that in the RMB-USD bilateral rate a 1% movement is indeed a very large shock. The standard deviation of the residuals from the first equation of the VAR - the exchange rate equation - is of 0.4%. Thus a 1% depreciation is more than a two standard deviation size change in the onshore rate.

Beyond the mode and the impact effects, we can observe that the confidence intervals around FX intervention narrow substantially after about 12 months from the initial shock. My estimates suggest that FX intervention falls to zero between 6 to 12 months after the initial policy shock. Thus, even though my model allows only for short-run inference, we have some confidence that policy induced FX purchases are short-lived. On the other hand, the effects of the policy shock on the exchange rate are very long lasting. FX intervention depreciates the onshore spot rate on impact and this depreciation persists for at least 2 years after the shock at the 68% confidence level. These confidence intervals for the exchange rate response, however, are extremely wide, especially upwards, and do not exclude continued depreciations that peak at 3.5/4% at the 15-20 month horizon after the initial shock. At the mode, the RMB-USD rate continues to depreciate for the first 10-12 months after the initial shock. This suggests that the effect of FX interventions persists even after these have ceased to continue.

Finally in the lowest panel of figure 1.3, I have plotted the IRF of the trade balance. I find that an exogenous purchase of 24.2 Bln USD has on impact a, statistically insignificant, negative effect in the order of -97.2 Mln USDs. Equivalently, for every 1 USD of FX intervention the trade balance deteriorates by less than 1 cent (0.01 USD). In the 68% highest likelihood set the model with the highest impact effect estimates that the trade balance improves only by ten cents (0.1 USD) for every dollar of FX intervention. Beyond the impact effect, the trade balance shows a small dip which is reversed within the first few months. The two opposing effects roughly cancel each other out. However, after this dynamic response, the IRF remains near zero for the remainder of the inference horizon. In sum, over the short-medium run the effect of such purchases on the trade balance is statistically insignificant and economically small. Another metric of the importance of FX intervention shocks for the trade balance can be garnered by looking at the variance decomposition. At the mode, the shock explains less than 5% of the variation in the trade balance. In the 90% highest likelihood set, the shock can explain at most 10% on impact and 20% over the other horizons of the total variation in the trade balance. While my model cannot sign the effect, it suggests that the quantitative relevance of this channel is rather marginal.

1.6.1 Robustness checks

The financial crisis

As discussed in section 1.2.2, a feature of the data that stands out from the managed float regime is the behavior of the RMB-USD rate between October 2008 and May 2010. In this period, the PBoC returned to peg the currency against the USD in response to the events of the financial crisis and the steep fall in global trade. There is a worry that the differing behavior of the exchange rate in this period is a result of a change in policy by the PBoC.

For this reason, I exclude the period between October 2008 and June 2010 from the managed float regime and re-estimate my model with this restricted sample.

The results are plotted in figure 1.4. The impact effect of the policy shock, however, remains practically unchanged. The USD purchases required to induce a 1% depreciation remain constant though the confidence interval widens a bit, especially downwards. This, however, is probably due to the fact that excluding the October 2008 to June 2010 period from the managed float period reduces my sample by about one third (102 observations to 76 observations). Thus, the precision of the estimate of the reduced falls. The dynamics of the FX intervention response and the spot rate, on the other hand, differ a little. At the mode, after 6 months of positive interventions the PBoC begins to sell FX assets and shortly thereafter the exchange rate begins to slowly appreciate. However, this appreciation is slow and the exchange rate remains depreciated for 2 years following the shock just as above.

The impact effect on the trade balance remains small - -1.71 Bln USD - especially when compared to the FX purchase of 24.2 Bln USD, though it is significantly bigger than the baseline. The estimates also suggest that the positive movement in the trade balance in the first few months is now statistically significant though extremely transitory, lasting 2 months at most. Once again, six months after the initial shock the effects on the trade balance are only very marginal. Looking at the variance decomposition across the two estimates also suggests that the impact of FX intervention on the trade balance is comparable. At the modal response it explains slightly less than 10% of the total variance in the trade balance across the entire inference horizon. As before, in the 90% highest likelihood set, the shock can explain at most 20% of the total variation in the trade balance.

A diffuse prior

A concern with any model identified exploiting structural change is that the change in the parameters is not limited to the subset of parameters *a priori* identified by the researcher. In my identification procedure, I assume that the variation in FX intervention attributable to fundamental shocks other than the policy shock - denoted with $g(\bullet)$ - is constant across exchange rate regimes. As discussed in section 1.3.3, there are reasons both in favour and against this parameter remaining constant. I therefore proceed to show that the fundamental inference that is made by my model does not differ much when introducing additional uncertainty around the value of $g(\bullet)$.

I relax assumption 4 in the following way. Let $g^{(1)}(\bullet)$ be the variation in FX intervention in the fixed exchange rate regime and $g^{(2)}(\bullet)$ that in the managed float period. Since both functions are scalar valued, there is no loss of generality by expressing their relationship as: $g^{(2)}(\bullet) = \varsigma g^{(1)}(\bullet)$. Since, $g^{(1)}(\bullet)$ is measured in the data, I can express my uncertainty over $g^{(2)}(\bullet)$ by letting ς be a random variable. The choice of prior (e.g. Uniform, Beta, etc) for ς could be justified by the researcher's prior over the likelihood of a large or small change in $g^{(i)}(\bullet)$ across periods. In this application, I work with a uniform distribution so that $\varsigma \sim U[\underline{\varsigma}, \overline{\varsigma}]$.

The primary concern in this paper is that an increase in the exchange rate flexibility induces an increase in speculative activity on the currency. I therefore assume that $\underline{\varsigma} = 1$, reflecting that the concern is around an increase of the USD inflow volatility, not a decrease. The choice of the upper bound is more arbitrary. In the results I present I set it to $\overline{\varsigma} = 5$. Thus, I allow the $g^{(2)}(\bullet)$ parameter to increase by up to 5 times, at any point of the posterior distribution. I rescale the entire posterior distribution of $g^{(1)}(\bullet)$ by this factor, not only the mean. In practice, I do this by taking a draw from the posterior distribution for $g^{(1)}(\bullet)$ and a draw from my prior over ς. Multiplying the two random variables we obtain a draw for $g^{(2)}(\bullet)$.

The results are plotted in figure 1.5. The first, obvious, effect of relaxing assumption 4 is that the uncertainty around the IRFs increases substantially. Note, however, that the modal response of all three variables is highly comparable to the baseline. The estimates in the middle panel suggest that the PBoC may need to purchase - at the upper bound - 80 Bln USD to induce a 1% depreciation versus 24.2 Bln in the baseline estimates. The effect of these purchases on the trade balance, though very comparable at the mode, is far more uncertain in USD terms since the confidence interval around the IRF in the first year widens significantly. However, also the average intervention size increases significantly. The variance decomposition still suggests that the explanatory power of the policy shocks for the total variation in the trade balance is small and the upper bound is consistent with the other two estimates.

1.7 Support for the identifying assumptions

In this section, I provide indirect evidence in favour of the identification scheme employed in my structural VAR analysis. I show that the IRF of the bilateral USD-RMB spot rate to the policy shock I have identified is consistent with IRF function identified using policy shocks identified at high-frequency.

1.7.1 Identification of the policy shocks

Martin (2013) identifies a set of 23 announcements referencing news regarding Chinese exchange rate policy and/or Chinese U.S. Treasury purchases. This set of events has two desirable properties that make identification of the policy shock relatively clean. The first property is that the set of events is selected using an exogenous rule to avoid sample selection. These events are collected either from announcements posted on the PBoC's website or by keyword search of news announcements published on the *Wall Street Journal*. For any event to be included in the final 23 a precise measure of the announcement time must also be included. The events are not restricted in any other way. Second these announcements are chosen to be explicitly exogenous and specific. They are specific because they are announcements about Chinese policy intentions made only by government officials. They are exogenous because they contain information about future policy and nothing else.

This last requirement is crucial. Identification based on high-frequency assumes that the change induced in the spot and forward rates in the event window must arise only from the change in the expected policy shock[31]. When this is not the case, identification fails. For example, if the announcement also releases information about other shocks hitting the economy then the change in the conditional expectation of the spot and forwards will vary both because of the announcement and because of the additional information released with the announcement. Thus the estimated coefficient from the surprises will conflate the two effects. The second requirement for identification is that the risk premium on the forward *not* change in the event window in response to the shock. In the sample period under analysis, the RMB-USD presents a statistically significant forward premium. This does not per se invalidate the inference so long as this premium is *constant* in the event window.

Though the set of events is carefully selected using the criteria described above, sample selection may be still a concern. In fact, differently from the FOMC meetings which follow a pre-established calendar, the set of announcements identified by Martin (2013) is set of instances in which PBoC authorities chose to divulge information. In the literature on FX interventions, there is some evidence that announced FX interventions are more effective at guiding the exchange rate than unannounced interventions (Fratzscher (2006)). These results are however based on countries with a free floating exchange rate. Nonetheless, it is possible that the economic shocks that are being compared in this analysis are different. Note, however that if the shocks were in fact different then this would suggest that the IRFs should differ.

Lastly, the set of events spans between July 21, 2005 and September 30, 2011. Prior to the end of fixed peg to the dollar there are no relevant announcements as Chinese officials never made any policy announcement regarding their exchange rate or foreign exchange policy.

[31]For more details see Section 3 of Faust et al. (2004).

1.7.2 Data on Non-Deliverable Forwards

In line with the literature, I use the change in basis points in the forwards to measure the change in the market's conditional expectations of the future exchange rate around the policy announcement. The peculiarity of China is that these over-the-counter forwards are traded offshore and are non-deliverable since the RMB is not convertible under the capital account. Even though these contracts are traded offshore, the underlying reference asset is the onshore RMB-USD spot rate. From Bloomberg, I obtain onshore spot and NDF end of day prices for the following maturities: 1M, 2M, 3M, 6M, 9M and 12M.

In the high-frequency literature, policy surprises are typically measured in short 30 minute to 2 hours windows using intradaily data to ensure clean identification of the policy shock. One of the limitations of the forwards data available from Bloomberg is that it is available only at daily frequency[32]. I, in part, circumvent this problem by exploiting the fact that Bloomberg reports in fact three different end of day prices. It reports the market[33] close price of the NDFs at the close of the Tokyo, London and the New York stock markets. These contracts are not traded in these venues; the closing times are just used for time reference purpose. Due to the different closing times (in local time) and the time zone difference between these three cities it is possible to construct surprise windows that are strictly smaller than the 24 hour windows between the NY close price pre-event and the NY close price post-event. Most events occur between the close of the NY stock exchange and the close of the Tokyo stock exchange reducing to about 13 hours the time between pre and post event prices.

1.7.3 Outlier analysis and results

Figure 1.6 graphs the scatter plots between the spot rate change and the change in the various forward contracts measured in basis points. As is easy to see, the sample is dominated

[32]It is available intradaily for the last 180 days. However, the intradaily quotes are bank specific rather than market quotes.

[33]NDF are traded over the counter by a number of financial institutions. Bloomberg uses an algorithm to aggregate individual quotes to form the market quote using a cutoff algorithm for quotes that have become stale.

by a large outlier that occurs on July 21st 2005 - the day of the policy change. Standard regression methods are well known to be heavily influenced by such outliers. As a first step then, I employ a robust regression method to identify the weight given to this observation by this procedure. Robust regression methods are a form of weighted regression, where the weights are chosen to guard against possible outliers and model misspecification.

I find that the robust regression assigns a weight of zero to that observation across all NDF maturities. This suggests that a robust estimate of the IRF should exclude the policy surprise that occurred on July 21st 2005. This is perhaps not surprising as in this particular policy surprise are embedded two pieces of news: (i) the change in regime - the announcement of a shift to a managed float, rather than a revaluation to another fixed peg, (ii) the policy surprise - the amount of revaluation in the bilateral rate. The second component is common with the rest of the sample but the first is not. For this reason I exclude this observation from the sample.

In table 1.3 I have reported the OLS estimates of the spot rate change on the change in different maturities of the forwards excluding the first observation. I wish to compare these estimates to those implied by my structural VAR and test that the predictions of the two models are consistent. From visual inspection of figure 1.3, one can see that the IRF identified at high-frequency - the dotted black line - does not differ very strongly from that identified in the structural VAR - the dark blue line. More formally, I carry out this test using a Wald test, with the null hypothesis that $\beta_{HF} - \beta_{SVAR} = 0$[34]. For this test to be correct, the asymptotic distribution of both the OLS estimates and the Bayesian VAR coefficients must be asymptotically normal. If both are normally distributed vectors, then the difference in the coefficients will still be normally distributed. Further, the variance-covariance matrix will simply be the sum of the two variance-covariance matrices. With 6 maturities of NDF, the test statistic will asymptotically follow a Chi-square distribution with 6 degrees of freedom. The Wald statistic has a value of 6.503, which implies that the

[34]Strictly speaking the test is ill defined. I improperly interpret the Bayesian posterior likelihood as a frequentist confidence interval when I conduct the Wald test.

null cannot be rejected at any standard confidence level[35]. This is unsurprising as inspection of figure 1.3 reveals that the two IRF estimates almost perfectly overlay each other. Thus eventough the power of the test is low, the almost identical point estimates suggest that the models do in fact give similar predictions in terms of the USD-RMB response to a policy shock.

1.8 Conclusion

In this paper, I have documented how under assumptions 1-5 it is possible to identify exogenous policy shocks in the form of FX intervention by the PBoC. To my knowledge, the identification scheme is a novel application of established identification methods (Kearns and Rigobon (2005), Bacchiocchi and Fanelli (2012)). I use these identified policy shocks to study the effect of FX intervention on the onshore RMB-USD spot and the trade balance. I find that purchases of FX assets depreciate the exchange rate and this effect is long lasting (1-2 years), especially when compared to the impact of FX intervention in countries with a free floating exchange rate regime. However, the scale of FX intervention the PBoC must undertake to affect the exchange rate is extremely large suggesting that the onshore USD supply is, in fact, rather elastic. Further, I show that these same FX shocks have economically small and statistically insignificant effects on the trade balance. At the mode, for every 1 USD of FX intervention the trade balance deteriorates by less than 1 cent (0.01 USD). In the 68% highest likelihood set the model with the highest impact effect estimates that the trade balance improves only by ten cents (0.1 USD) for every dollar of FX intervention. In the set of likely models, this estimate is the estimate which is most favourable to the neo-mercantilistic hypothesis. Nonetheless, even in this case, I find that the effect is small. Over longer horizons, I find that the per dollar effect of FX intervention on the trade balance does not differ much from the above estimates. When I relax the assumption that the variance of non-policy shocks be constant across regimes, I once again find evidence that this effect is at the mode near zero and in the most favourable model it remains small. Overall, my estimates suggest that FX intervention doesn't affect the trade balance, and if

[35]The relevant threshold values are 10.64, 12.53, 16.81 to reject the null at the 10%, 5% and 1% level respectively.

it does this effect is small. Finally, I employ the policy shocks identified by Martin (2013) to provide indirect evidence of my model's predictions. I find that the IRF of the bilateral RMB-USD rate identified by his policy shocks is consistent with the IRF of my structural VAR. Though this does not constitute a formal test of my model's identifying assumptions, I interpret this as supporting evidence for the plausibility of my inference.

1.9 Appendix

Proof of identification

From the system in (1.11), the reduced form 2×2 covariance matrix, Σ_S, may be derived to be composed of the following elements[36]:

$$\Sigma_{1,1} = (\frac{1}{|A0|})^2 \left\{(1-\beta_2\gamma_2)^2\sigma_D^2 + \alpha_1^2\sigma_1^2 + \alpha_1^2\beta_2^2\sigma_2^2 + \alpha_1^2 2\beta_2\sigma_{1,2}\right\} \tag{1.26}$$

$$\Sigma_{1,2} = (\frac{1}{|A0|})^2 \left\{(1-\beta_2\gamma_2)(\beta_2\gamma_1+\beta_1)\sigma_D^2 + \alpha_1\sigma_1^2 + \alpha_1\beta_2^2\sigma_2^2 + \alpha_1 2\beta_2\sigma_{1,2}\right\} \tag{1.27}$$

$$\Sigma_{2,2} = (\frac{1}{|A0|})^2 \left\{(\beta_2\gamma_1+\beta_1)^2\sigma_D^2 + \sigma_1^2 + \beta_2^2\sigma_2^2 + 2\beta_2\sigma_{1,2}\right\} \tag{1.28}$$

where

$$\frac{1}{|A_0|} = \frac{1}{(1-\alpha_1\beta_2\gamma_1) - (\gamma_2\beta_2 + \beta_1\alpha_1)} \tag{1.29}$$

$$= \frac{1}{1-\beta_2\gamma_2 - \alpha_1(\beta_2\gamma_1+\beta_1)} \tag{1.30}$$

$$= \frac{\frac{1}{1-\beta_2\gamma_2}}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})} \tag{1.31}$$

Define the scalar valued functions $g(\bullet)$ and $h(\bullet)$ to be:

$$g(\bullet) = \frac{\sigma_1^2 + \beta_2^2\sigma_2^2 + 2\beta_2\sigma_{1,2}}{(1-\beta_2\gamma_2)^2} \tag{1.32}$$

$$h(\bullet) = \frac{\beta_1 + \beta_2\gamma_1}{1-\beta_2\gamma_2} \tag{1.33}$$

Consider (1.26):

$$\Sigma_{1,1} = (\frac{\frac{1}{1-\beta_2\gamma_2}}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})})^2 \left\{(1-\beta_2\gamma_2)^2\sigma_D^2 + \alpha_1^2\sigma_1^2 + \alpha_1^2\beta_2^2\sigma_2^2 + \alpha_1^2 2\beta_2\sigma_{1,2}\right\} \tag{1.34}$$

$$= (\frac{1}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})})^2 \left\{\sigma_D^2 + \alpha_1^2\frac{\sigma_1^2 + \beta_2^2\sigma_2^2 + \alpha_1^2 2\beta_2\sigma_{1,2}}{(1-\beta_2\gamma_2)^2}\right\} \tag{1.35}$$

$$= (\frac{1}{1-\alpha_1 h(\bullet)})^2 \left\{\sigma_D^2 + \alpha_1^2 g(\bullet)\right\} \tag{1.36}$$

[36]I omit the dependence on the exchange rate regime to simplify the notation

Consider (1.27):

$$\Sigma_{1,2} = (\frac{\frac{1}{1-\beta_2\gamma_2}}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})})^2 \left\{(1-\beta_2\gamma_2)(\beta_2\gamma_1+\beta_1)\sigma_D^2 + \alpha_1\sigma_1^2 + \alpha_1\beta_2^2\sigma_2^2 + \alpha_1 2\beta_2\sigma_{1,2}\right\} \tag{1.37}$$

$$= (\frac{1}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})})^2 \left\{\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2}\sigma_D^2 + \alpha_1\frac{\sigma_1^2+\beta_2^2\sigma_2^2+2\beta_2\sigma_{1,2}}{(1-\beta_2\gamma_2)^2}\right\} \tag{1.38}$$

$$= (\frac{1}{1-\alpha_1 h(\bullet)})^2 \left\{h(\bullet)\sigma_D^2 + \alpha_1 g(\bullet)\right\} \tag{1.39}$$

Consider (1.28):

$$\Sigma_{2,2} = (\frac{\frac{1}{1-\beta_2\gamma_2}}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})})^2 \left\{(\beta_2\gamma_1+\beta_1)^2\sigma_D^2 + \sigma_1^2 + \beta_2^2\sigma_2^2 + 2\beta_2\sigma_{1,2}\right\} \tag{1.40}$$

$$= (\frac{1}{1-\alpha_1(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})})^2 \left\{(\frac{\beta_2\gamma_1+\beta_1}{1-\beta_2\gamma_2})^2\sigma_D^2 + \frac{\sigma_1^2+\beta_2^2\sigma_2^2+2\beta_2\sigma_{1,2}}{(1-\beta_2\gamma_2)^2}\right\} \tag{1.41}$$

$$= (\frac{1}{1-\alpha_1 h(\bullet)})^2 \left\{(h(\bullet))^2\sigma_D^2 + g(\bullet)\right\} \tag{1.42}$$

Thus Σ_S may be expressed as:

$$\Sigma_S = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{bmatrix} \tag{1.43}$$

$$= (\frac{1}{1-\alpha_1 h(\bullet)})^2 \begin{bmatrix} \sigma_D^2 + \alpha_1^2 g(\bullet) & h(\bullet)\sigma_D^2 + \alpha_1 g(\bullet) \\ h(\bullet)\sigma_D^2 + \alpha_1 g(\bullet) & (h(\bullet))^2\sigma_D^2 + g(\bullet) \end{bmatrix} \tag{1.44}$$

which follows from (1.36), (1.39) and (1.42).

Further in the fixed exchange rate period the submatrix Σ_S:

$$\Sigma_S = (\frac{1}{1-\alpha_1 h(\bullet)})^2 \begin{bmatrix} \sigma_D^2 + \alpha_1^2 g(\bullet) & h(\bullet)\sigma_D^2 + \alpha_1 g(\bullet) \\ h(\bullet)\sigma_D^2 + \alpha_1 g(\bullet) & (h(\bullet))^2\sigma_D^2 + g(\bullet) \end{bmatrix} \tag{1.45}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & g(\bullet) \end{bmatrix} \tag{1.46}$$

since $\alpha_1 = 0$ and $\sigma_D = 0$. Thus the variance of FX intervention directly identifies $g(\bullet)$.

Figures

Figure 1.1: RMB-USD exchange rate regimes

In blue, I have plotted the time-series of the onshore RMB-USD rate measured on the last day of each month. Sample begins in January 1998 and ends in December 2013. The fixed exchange rate period spans the period January 1998 to June 2005. The managed float regime spans the period July 2005 to December 2013. The vertical red line separates the fixed exchange rate regime from the managed float period. Between October 2008 and June 2010, the PBoC effectively re-pegs the exchange rate to the USD. The period in question is highlighted by the two vertical black lines.

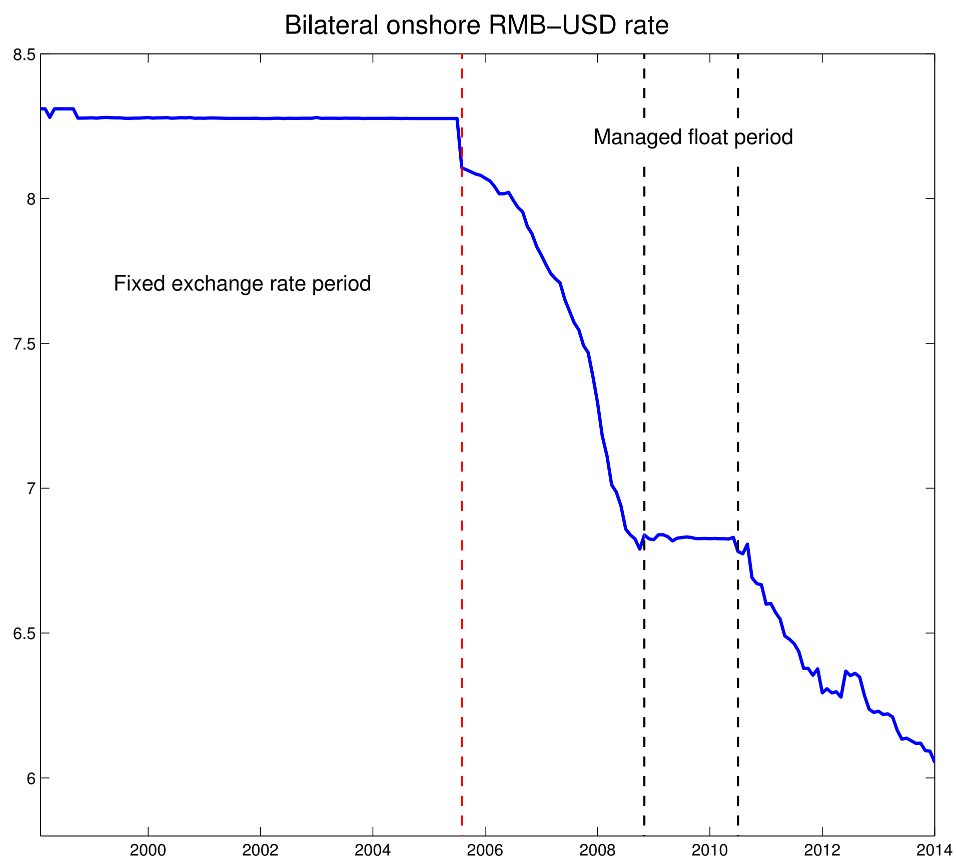


Figure 1.2: Estimate of FX intervention time-series

The figure plots the estimated latent process I_t (in dark blue) along with the 68% standard errors bands (in red). The time-series measures in USD Billions the purchases of FX assets by the PBoC. The measure of I_t is the output of a mixed frequency Kalman filtering procedure. The mean and standard deviation of the latent state are based on 10,000 draws from the posterior distribution. The monthly signal, given by the change in FX stocks (ΔFX_t), is plotted in green. The quarterly signal, given by the flow of FX reserves ($FXBOP_t$), is absent from the plot.

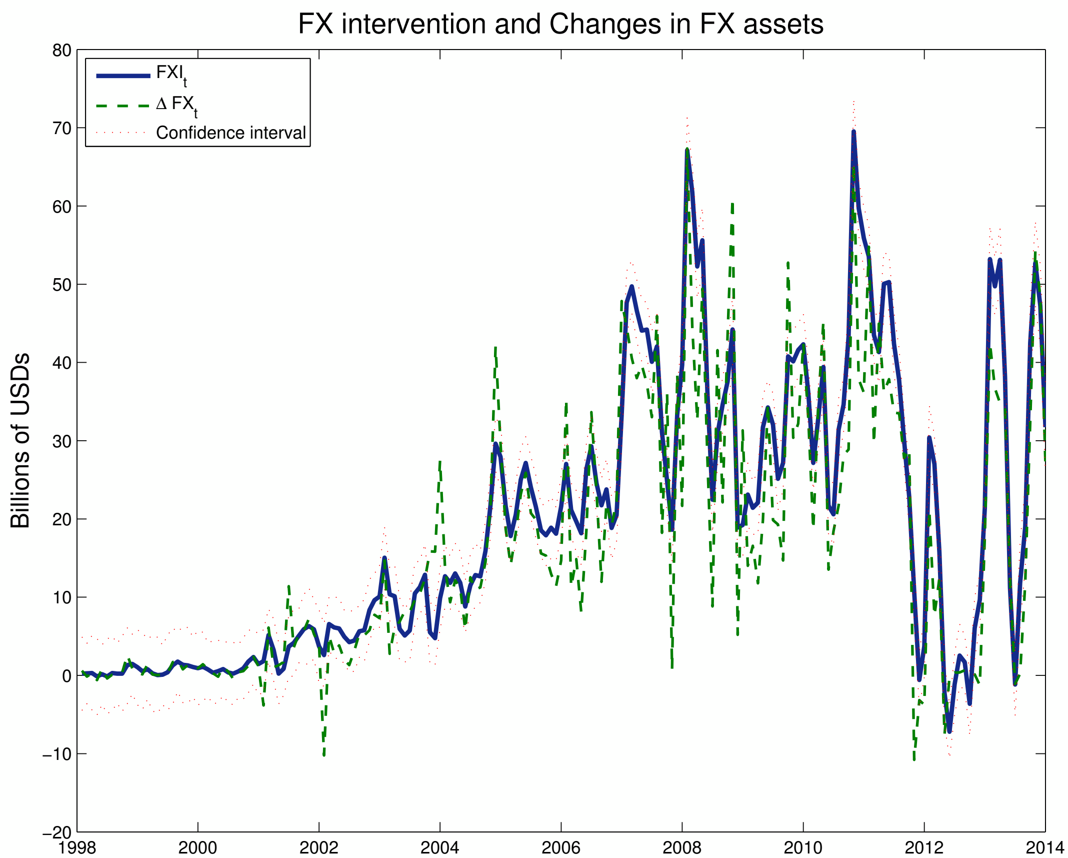


Figure 1.3: Impulse Reponse Functions: Baseline results

The figure plots the IRFs of the RMB-USD rate, FX intervention and trade balance in response to a policy shock. The responses trace the response of each variable to a policy shock in the managed float regime. The policy shock is normalized to induce a 1% depreciation of the RMB-USD spot. The bold blue lines are the modal model. The dotted light (dark) red line is the outer contour of the 68% (90%) confidence intervals.

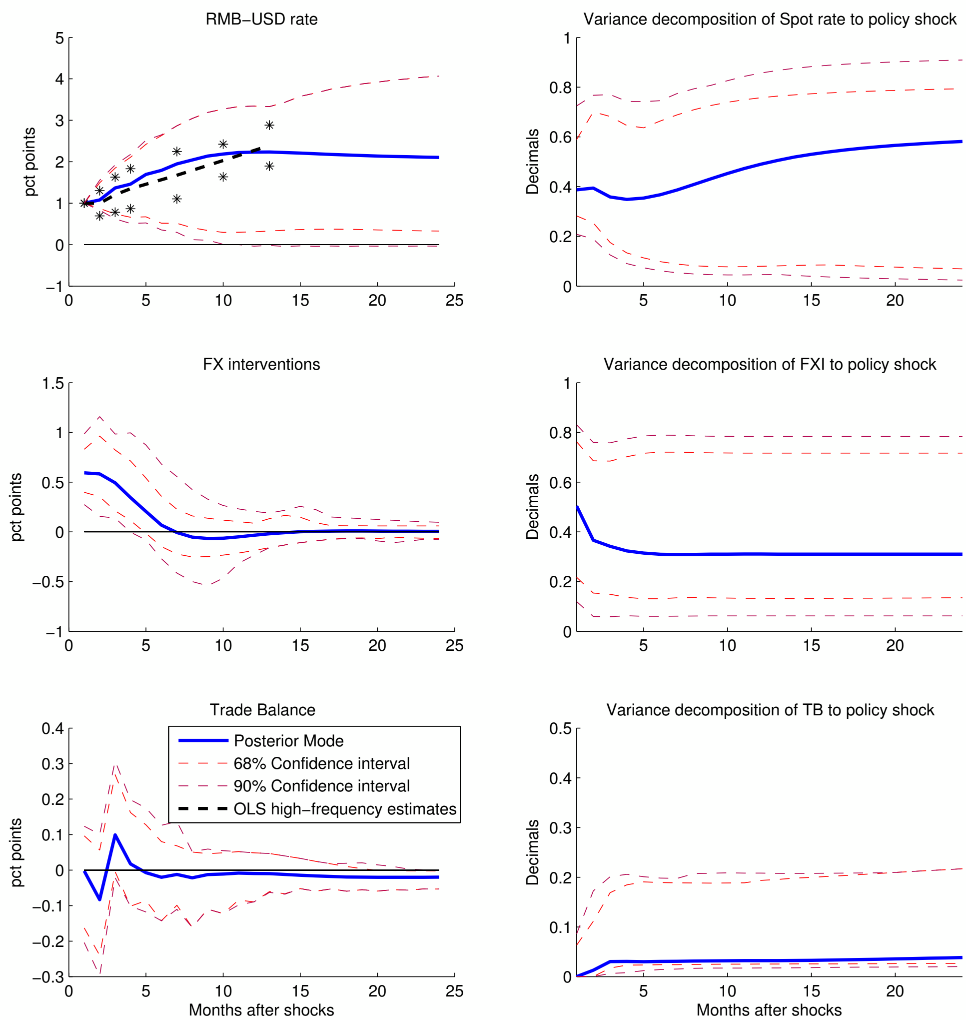


Figure 1.4: Impulse Reponse Functions: Excluding October 2008-July 2010

The figure plots the IRFs of the RMB-USD rate, FX intervention and trade balance in response to a policy shock. The responses trace the response of each variable to a policy shock in the managed float regime. In this computation, the period between October 2008 and July 2010 is excluded from the managed float sample. The policy shock is normalized to induce a 1% depreciation of the RMB-USD spot. The bold blue lines are the modal model. The dotted light (dark) red line is the outer contour of the 68% (90%) confidence intervals.

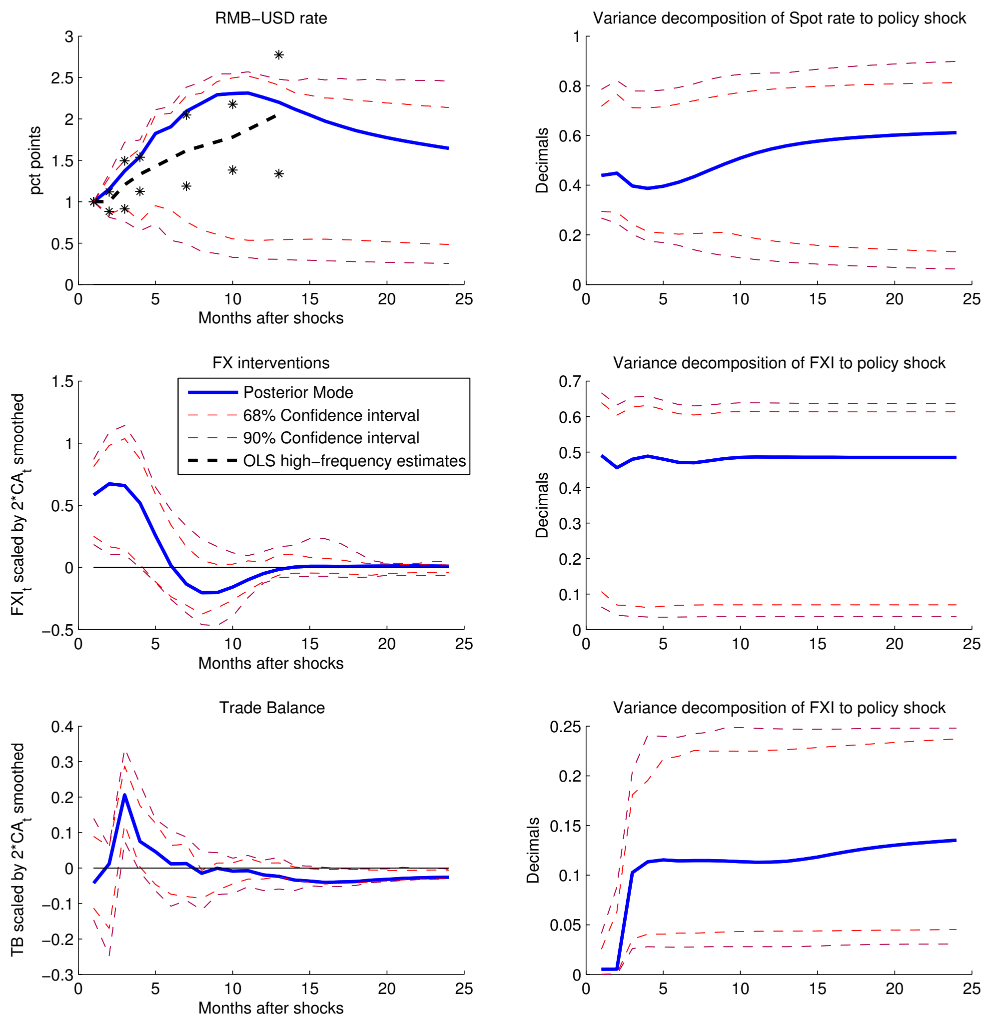


Figure 1.5: Impulse Reponse Functions: Diffuse prior

The figure plots the IRFs of the RMB-USD rate, FX intervention and trade balance in response to a policy shock. The responses trace the response of each variable to a policy shock in the managed float regime. In this computation, assumption 2 is relaxed. It is assumed that the period 2 value of the variance - $g(\Phi_{NP})$ - may take any value equal to or up to five times larger than the period 1 estimate. Once again, the policy shock is normalized to induce a 1% depreciation of the RMB-USD spot. The bold blue lines are the modal model. The dotted light (dark) red line is the outer contour of the 68% (90%) confidence intervals.

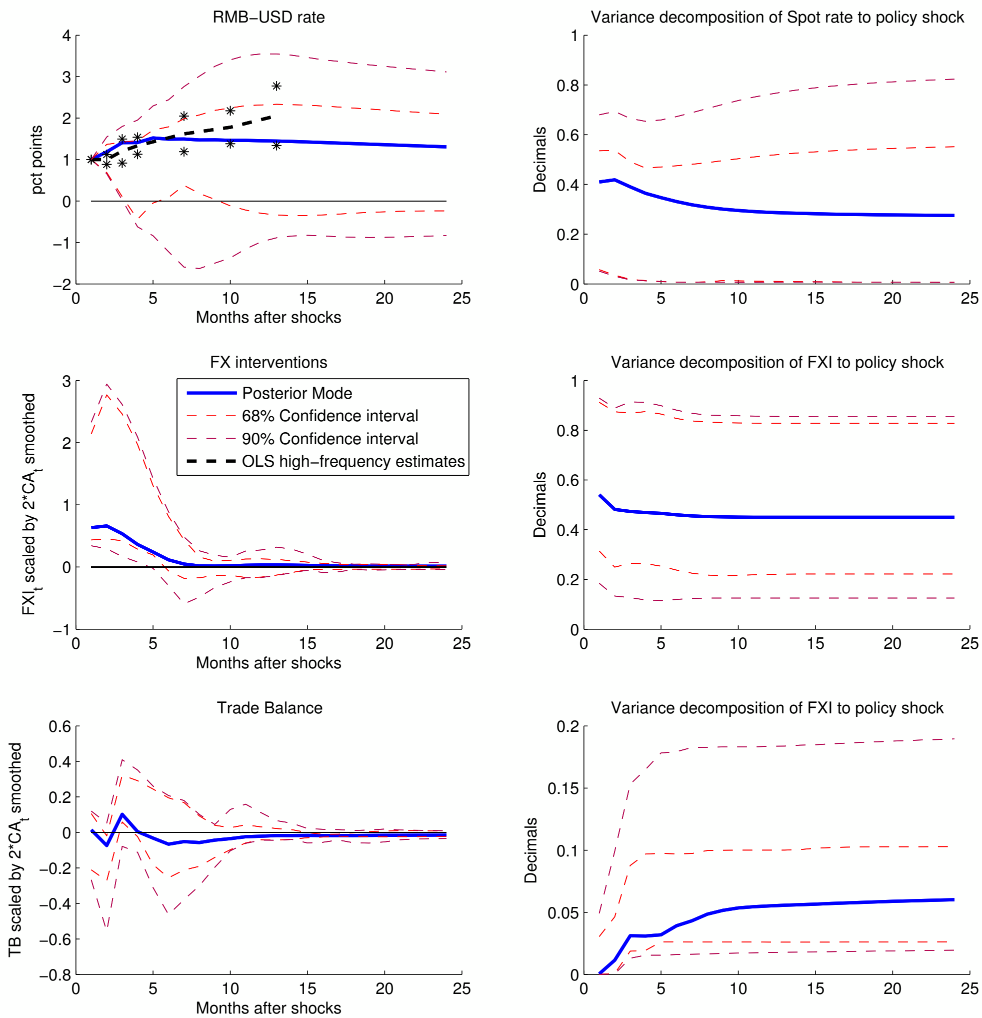
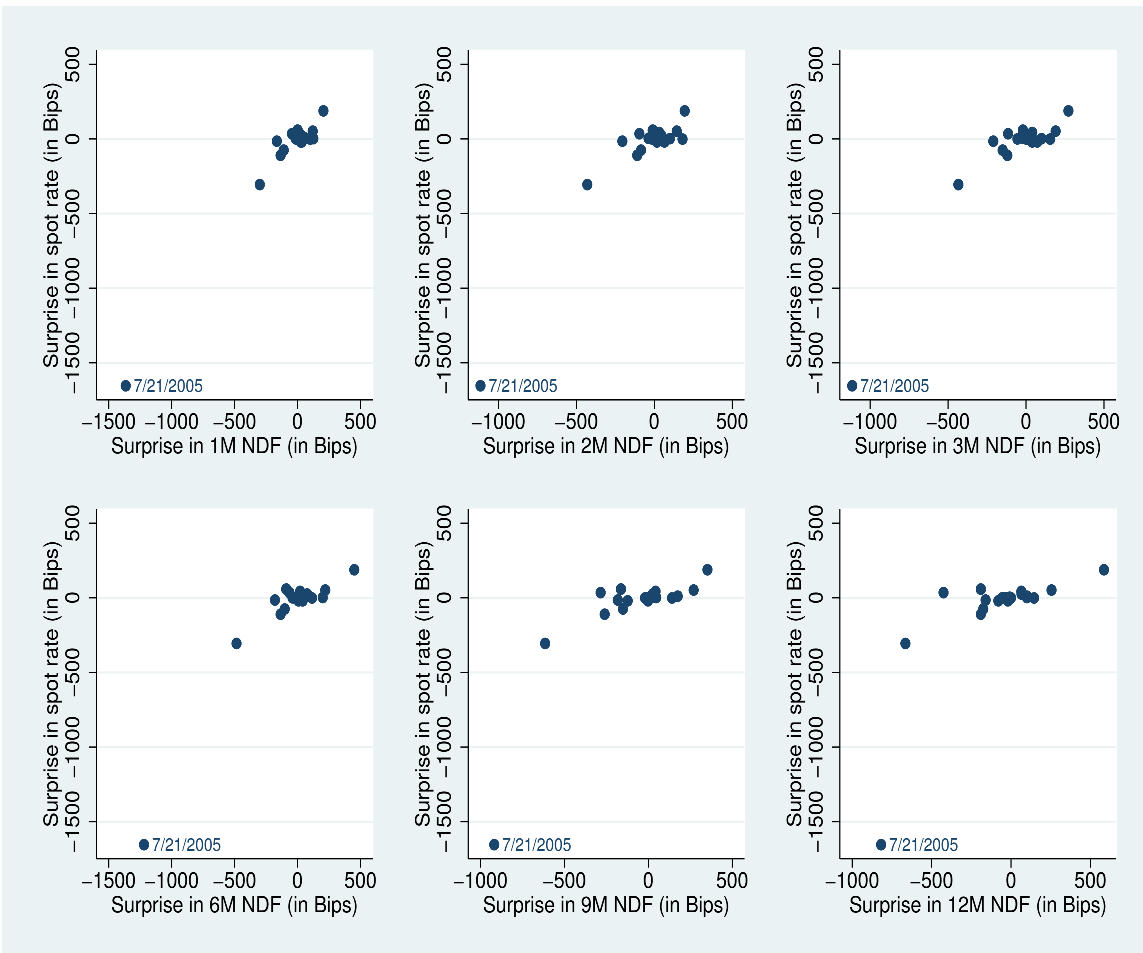


Figure 1.6: Changes in the spot and forward RMB-USD rates in the event windows

The figure displays the scatter plots of the change in the RMB-USD spot rate against the change in the Non-Deliverable Forward (NDF) rate in basis points in the policy surprise windows identified by Martin (2013). The suprise occuring on July 21st, 2005 is labelled "7/21/2005" and clearly appears to be an outlier in all panels.



Tables

Table 1.1: Priors and Posteriors for currency shares

The table lists priors (left panel) and posteriors (right panel) for the various value shares of the currencies used to model the valuation effect (see (1.25)). Priors are based on the work of Sheng (2013). Posterior mean and standard deviation are based on 10,000 draws from the posterior distribution of the currency shares after an initial burn-in of 5,000 draws.

	Priors		Posteriors	
Coef	Mean	Std Dev	Mean	Std Dev
β_1^{Euro}	5%	1%	5.8%	1.2%
β_2^{Euro}	23%	2%	24.6%	2.3%
β^{Yen}	2.7%	1%	4.2%	1.1%
β^{AUD}	3%	1%	1.7%	1%
β^{GBP}	3%	1%	4.6%	0.8%

Table 1.2: Recapitalizations undertaken by the PBOC

The table lists recapitalizations undertaken by the PBoC using FX assets under control of the State Administration for Foreign Assets (SAFE). The table is based on data collected by Zhang (2010), Table 1 pg. 39.

Institution	Date	Bn (USD)
Bank of China	Dec 2003	22.5
China Construction Bank	Dec 2003	20
Jianyin Investment Company	Dec 2003	2.5
Industrial and Commercial Bank of China	April 2005	15
China reinsurance (group) Co.	April 2007	2
National Development Bank	Dec 2007	20
Agricultural Bank of China	Oct 2008	19

Table 1.3: High-Frequency based estimates of RMB-USD IRF

The table lists the point estimates and standard errors of the 6 separate univariate regressions of the spot rate change against the change in one of the 6 forward maturities in the surprise window. These estimates exclude from the sample the first policy surprise of July 21st, 2005 for the reasons discussed in the paper.

SPOT	1M NDF	2M NDF	3M NDF	6M NDF	9M NDF	12M NDF
β	1.001***	1.205***	1.332***	1.618***	1.780***	2.057***
σ_β	(0.154)	(0.208)	(0.219)	(0.260)	(0.334)	(0.409)
Observations	22	22	22	22	22	22
R-squared	0.679	0.627	0.650	0.659	0.587	0.558

CHAPTER 2

Gross Bank Credit Flows: Evidence from the unit banking system in the state of New York, 1912-1932 - with Haelim Park

2.1 Introduction

In recent years, a number of papers have documented interesting and independent dynamics of gross bank credit flows that underlie net bank credit flows. In fact, aggregate changes in net credit almost always mask the considerable heterogeneity in changes in bank credit across banks. Scholars have measured this heterogeneity in terms of gross creation and destruction of credit and the rate at which credit is reallocated across banks above and beyond the net change. This excess credit reallocation is considered a proxy of credit market frictions as it provides a measure of the additional reallocation of credit that must occur to accommodate changes in net credit.

Several empirical papers have shown that large gross credit flows exist in the U.S. banking system. In particular, Dell'Arriccia and Garibaldi (2005), Craig and Haubrich (2013), and Contessi and Francis (2010) construct measures of gross credit flows using the Call Reports for the U.S. banking system. They find that credit creation and destruction coexist over the business cycle and that excess credit reallocation is countercyclical as a result of a decline in the creation margin and a rise in the destruction margin during recessions. A related literature studies credit reallocation across non-financial firms. Herrera et al. (2011) construct such measures using firm-level financial data from the U.S. Compustat tapes and documents. They conclude that excess credit reallocation is pro-cyclical due to a rapid decline in the creation margin with a moderate rise in the destruction margin.

Our empirical work is an extension of this line of research. In particular, we document stylized facts about gross bank credit flows in New York State between 1912 and 1932. To do so, we construct a bank level dataset from the reports of the superintendents of the banking department, which included all state-chartered banks and trust companies. This dataset presents two majors advantages. First, the dataset is composed mostly of unit banks[1]. This helps alleviate aggregation bias that arises in modern data since a unit bank serves the local market and does not have any connecting banks (i.e. branches) in other areas. Second, this dataset is characterized by loan categories that are structured in function of the collateral that secures the loan. This allows us to study how the creation and destruction margins vary in function of the underlying collateral of a loan.

In line with the literature, we find that credit creation and destruction coexist across the business cycle. However, we find that mean levels of creation, destruction and excess reallocation are higher than those previously documented. In particular, we find that excess reallocation of credit per quarter is in the order of 4.71%, rather than 2.6%. Thus, the heterogeneity in behaviour and the potential cost of credit frictions is almost twice that previously estimated. We also find that excess reallocation is strongly pro-cyclical, rather than countercyclical as documented by the literature. We suggest that this may be due to differences in the cyclical behaviour of bank deposit flows across the business cycle before and after the institution of deposit insurance (Pennacchi (2006)).

In addition, we document how the behaviour of gross credit flows varies in function of the type of collateral underlying the loan. In particular, we show that collateralized loans

[1] The unit structure of the banking system in the U.S. contributed to inefficient allocation of credit. The unit banking system created barriers to entry, which prevented productive competition amongst banks, especially in rural areas. In addition, it produced lack of diversification of loan risk within banks as each bank's portfolio risk reflected the operations of its local economy. In agricultural areas, the income of banks was closely tied to changes in the prices and harvests of one or two crops. The unit banking structure hindered financial integration across regions, which resulted in large differences in interest rates across regions since banks can easily move funds across regions to accommodate differences in demand. Lastly, unit banking caused a growing mismatch between the supply and demand of credit. Small banks could not lend the sums needed by large industrial firms. When the scale of industry grew substantially in the nineteenth century, banks could not keep up with its demand. As a result, industries replaced bank loans with other means of financing by the end of the nineteenth century.

are pro-cyclical in the creation margin, but acyclical in the destruction margin. On the other hand, uncollateralized loans are acyclical in creation but countercyclical in destruction. As well, we find that creation is relatively more volatile than destruction for collateralized loans, while the opposite is true for uncollateralized loans. These results suggest that collateral reduces the informational asymmetry between borrowers and lenders making the creation margin more volatile and responsive to business cycle fluctuations than it would otherwise be. However, rather than attenuating the asymmetry between credit creation and destruction, collateral reverses this asymmetry. A mechanism by which collateral aids the formation of credit relationships but impedes their separation rationalizes our empirical results.

Our paper improves upon earlier studies in three main dimensions. First, our study provides more accurate measures of gross credit creation and destruction. This occurs due to the unit banking structure we study. In fact, since loan creation and destruction may not be contemporaneously observed within the smallest unit of observation, aggregation seriously affects our ability to measure such heterogeneity within a banking conglomerate. Second, we contribute to the literature by examining the role of collateral. A number of studies find that collateral reduces informational asymmetries (Bester (1987)), by reducing credit risk and moral hazard, thus increasing the ease with banks may expand credit. Specifically, we study the properties of gross credit flows in function of the collateral underlying the loans. This is particularly of interest because the asymmetry between the gross creation and destruction margins in this literature arises because the creation of loans is subjects to information frictions but the destruction process is not (Dell'Arriccia and Garibaldi (1998)). Third, the unit banking structure is advantageous because banking activities are extremely local. This allows us to identify regional shocks far more precisely.

The rest of the paper is organized as follows. Section 2.2 discusses the data we employ. Section 2.3 describes the methodology. Section 2.4 describes stylized facts about gross credit flows in the state of New York. Section 2.5 concludes.

2.2 Data

To characterize aggregate gross credit flows we require two distinct types of information. First, we require information on loans outstanding at the bank level. Second, to appropriately characterize the cross-section of changes in credit, we must properly account for any merger activity. We discuss each dataset in turn.

We collect and computerize bank level balance sheets from the annual report of the superintendent of the banking department of the state of New York. These annual reports published balance sheet information for all state chartered banks and trust companies in the state of New York. Data on nationally chartered banks, however, are not included in this report[2]. The report underwent major changes in 1911 and in 1939. In 1911, it started publishing balance sheet information of trust companies at the quarterly frequency[3]. In 1939, the annual report began to publish balance sheet categories that were inconsistent with those of previous reports. In addition, the publication of balance sheet statements was discontinued between 1933 and 1934. Given the limitations of the data, we restrict our sample period to 1912-1932. In this period, reporting was done on a quarterly basis, the structure of balance sheets is comparable across time and there are no gaps in the time-series of surviving banks.

Figure 2.1 displays the balance sheet categories of a state commercial bank in this period. For this study, we focus on four types of asset categories that represent credit. The first category is mortgages owned, which are loans on secured real estate[4]. Banks were allowed to make mortgage loans on farmland within one hundred miles of the city, though these mortgage loans were accompanied by several restrictions. The second category is loans and discounts secured by bond and mortgage, deed and other real estate collateral.

[2]As well, data on nationally chartered banks is not included in our dataset. This is because there is a large discrepancy between reports for state banks and national banks. The OCC reported balance sheets for national banks once a year and provided different asset categories.

[3]Prior to this change, these balance sheets were published at yearly frequency. This change in regulation originated in the desire to regulate trust companies which had been responsible for causing the panic of 1907.

[4]National banks were prevented from making mortgage loans until the passage of the Federal Reserve Act.

However, a deed was taken as a mortgage, but not an absolute transfer of ownership. Moreover, well-managed banks avoided deeds. The third category is loans and discounts secured by other collateral. These are loans secured by anything except for real estate security, Liberty Bonds, stocks and bonds listed on the stock exchanges, and unlisted securities. The fourth category is loans, discounts, and bills purchased not secured by collateral. These are loans represented by promissory notes.

We also collect information on entry, exit, mergers, and changes in charter status of state chartered financial institutions[5]. While previous studies were concerned with spurious gross credit flows due to merger activities, they were not as concerned about the entry and exit of banks. In contrast, our study may suffer from spurious gross credit flows due to the entry and exit of banks in addition to merger activities. Due to the absence of national banks in our dataset, we cannot presume that entering (exiting) banks are new (closing) banks. This is because new banks appear in our dataset when nationally chartered banks enter (exit) our sample by switching their charters or by merging with a state chartered bank. We discuss how we deal with this issue in section 2.3 and its effect on our estimates of aggregate credit flows.

New York is an ideal state to study gross credit flows for two reasons. First, the state represented a large share of total loans in the United States. During our sample period, national and state chartered banks in New York held an average of over 21 percent of all loans in the United States. In addition, state chartered banks played an important role in the state banking industry for they represented about 60 percent of total loans in New York State. Second, New York State established a banking system with a great diversity. New York's banking industry was made up of banks of all sizes and types, from large money centers and global banks in New York City, to medium-sized banks with active manufacturing and industrial bases, to small banks in rural areas.

[5]The annual report publishes merger, charter switch, name change, and exit dates for of all state banks and trusts which it oversees. Establishment dates for each bank are instead collected from Rand-McNally Banker's directory.

There are a number of important features of the sample we study. First, we analyze a period that is characterized by a prolonged lending boom (1920s) and a subsequent severe bust. This was reflected in a concurrent expansion of both the number of banks and loans outstanding in the period 1912-1929 and their contraction in the bust period following 1929. Relative to the number of banks, total banking assets grew far more rapidly. In fact, outstanding loans increased almost six-fold between 1912 and 1929. This run up in credit, however, was quickly reversed as can be seen in table 2.1. This credit boom was roughly similar in magnitude both in New York City and the New York State. However, in New York City the ramp up and the fall were both sharper. Second, in our sample there is substantial inequality in the size of the loan book as measured by the Gini index. However, both small and big banks experienced credit expansion in the 1920s as the Gini index did not increase very rapidly between 1912 and 1929. Third, the average level of banking concentration in our sample is very low. This is in large part due to the unit banking structure of the financial system. In fact, especially outside of New York City, banks were very small and extremely tied to the local market[6]. Across time, we observe that the fall in the number of banks is accompanied by an increase in banking concentration following the 1929 bust. In fact, the bust triggered a number of mergers that consolidated the banking sector. This consolidation wave was particularly strong in New York City.

2.3 Methodology

Following Dell'Arriccia and Garibaldi (2005) and Craig and Haubrich (2013), we decompose net credit growth into gross credit creation and gross credit destruction. This approach faces five main methodological issues. First, it may underestimate gross flows since we do not observe the simultaneous creation and destruction of loans. Second, it may underestimate gross flows if balance sheets are available at the institutional level rather than at the branch level. Third, it may overestimate credit flows due to loan trading amongst banks. Fourth, the interpretation of credit creation (destruction) may be unclear since we

[6] Together these two facts rationalize how there can be high levels of size inequality but low levels of banking concentration. In fact, our sample is composed in large part by small banks and a few big banks. These, however, do not constitute the majority of credit outstanding since the Herfindal index is low.

cannot distinguish whether credit creation occurs as a result of the extension of credit to new clients or the increase in their existing stocks of loans. Lastly, it may misestimate credit flows due to merger activity.

The use of historical data alleviates concerns arising from two methodological issues. First, it allows us to suffer less from the underestimation of gross flows due to the unit banking structure. In fact our balance sheets, effectively, represent branch level information rather than institution level information due to the lack of branching[7]. Second, it allows us to suffer less from the overestimation of gross flows due to loan trading. Due to the high level of asymmetric information, lending was very relationship based and local in this historical period (Bernanke (1983), Petersen and Rajan (2002)). In consequence, loan trading was very uncommon if not completely absent. This implies that changes in loans outstanding arise entirely due to changes in the overall supply of credit to firms, rather than credit being reallocated across banks alone.

At the same time, the use of our historical dataset introduces three different types of new issues. First, we may misestimate gross flows since we exclude nationally chartered banks. Second, it worsens the bias introduced by mergers since we cannot accurately measure credit creation and destruction when state banks merge with national banks. Third, it may misestimate gross flows if we do not control for spurious entries and exits of banks in and out of our sample caused by changes in charter status. The latter two issues are related because in both instances the growth rate of credit in the merger or charter switch period will be unobserved.

To characterize gross credit flows we compute for each bank in our sample the change in credit, $\Delta l_{i,t}$, as the difference in credit outstanding at t and $t-1$. When two banks merge, however, this approach is not feasible since one of the banks will have ceased to exist. We therefore employ the methodology of Dell'Arriccia and Garibaldi to estimate credit growth in the event of a merger. We refer to this as the adjusted change in credit and it is defined

[7]The exception is New York City in which branching within the five city boroughs was allowed.

below.

Consider a merger between bank i (surviving bank) and bank j (non-surviving bank) at time t. The adjusted change in credit for bank i is defined as:

$$\Delta\tilde{l}_{i,t} = \Delta l_{i,t} - \phi_{i,j}(t) l_{j,t-1} \tag{2.1}$$

where $l_{i,t}$ is the value of loans outstanding at bank i at t, $\Delta l_{i,t}$ is the change in loans outstanding at bank i and $\phi_{i,j}(t)$ is an indicator function that takes the value of 1 if bank i acquires bank j at t. The difference between the unadjusted change in credit, $\Delta l_{i,t}$, and the adjusted change in credit, $\Delta\tilde{l}_{i,t}$, is that the latter accounts for the fact that part of the credit growth experienced by bank i is determined by the absorption of bank j. Thus, credit creation in the combined entity is the total change in credit for bank i, the surviving bank, minus the previous period's value of loans that it inherits from the acquired bank.

The difficulty introduced by our dataset is that when a state and national bank merge, the value of (2.1) is not defined. To see this, consider a state bank i (surviving bank) merging with a national bank j (non-surviving bank). Since bank j is a nationally chartered bank, the value of $l_{j,t-1}$ is unknown since our dataset only covers state chartered banks. Thus, the adjusted change in credit cannot be computed and remains undefined. On the other hand, the unadjusted change in credit may be computed, however, it will systematically overestimate $\Delta\tilde{l}_{i,t}$.

Thus in our sample there are, in fact, two types of mergers: mergers between state chartered banks and mergers between a state and a nationally chartered bank. The difference is of consequence because, given our data, we may correctly characterize changes in credit for mergers between state banks but not for those between a state and national bank. In these instances the adjusted credit growth may not be computed and is undefined in the quarter in which the merger occurs. The unadjusted credit growth may be computed, however, this figure systematically overestimates credit creation (or destruction). Changes in the charter status of banks introduce an identical problem. In fact, when a nationally chartered bank switches to a state charter we are also unable to compute the growth in credit for this institution since we do not observe its balance sheet in the previous period. Treating this bank as a new bank generates spurious credit creation. Thus, for this subset of banks, either the change in credit is unobserved or it is systematically overestimated.

To understand how this issue impacts our measures of aggregate credit reallocation, we proceed to define these quantities. Gross creation and destruction rates are computed by aggregating the cross-section of positive and negative changes in credit respectively and both are scaled by the total value of loans outstanding in the previous period:

$$POS_t = \frac{\sum_i \Delta\tilde{l}_{i,t}}{\sum_i l_{i,t-1}} \qquad \text{for bank } i \text{ s.t } \Delta\tilde{l}_{i,t} \geq 0 \tag{2.2}$$

$$NEG_t = \frac{\sum_i |\Delta\tilde{l}_{i,t}|}{\sum_i l_{i,t-1}} \qquad \text{for bank } i \text{ s.t } \Delta\tilde{l}_{i,t} < 0 \tag{2.3}$$

these quantities may be also expressed as[8]:

$$POS_t = \frac{\sum_{i|\Delta\tilde{l}_{i,t}>0} \Delta\tilde{l}_{i,t}}{\sum_i l_{i,t-1}} = \sum_{i|g_{i,t}>0} w_{i,t} g_{i,t} \tag{2.4}$$

$$NEG_t = \frac{\sum_{i|\Delta\tilde{l}_{i,t}<0} \Delta\tilde{l}_{i,t}}{\sum_i l_{i,t-1}} = \sum_{i|g_{i,t}<0} w_{i,t} |g_{i,t}| \tag{2.5}$$

where

$$g_{i,t} = \frac{\Delta\tilde{l}_{i,t}}{0.5(l_{i,t-1} + l_{i,t})} \tag{2.6}$$

$$w_{i,t} = \frac{0.5(l_{i,t-1} + l_{i,t})}{\sum_i l_{i,t-1}} \tag{2.7}$$

The term $g_{i,t}$ is the growth rate of credit at bank i[9][10]. The term $w_{i,t}$ is the weight of $g_{i,t}$ in the cross-section of credit growth. It is the average size of bank i between $t-1$ and t

[8]See Dell'Arriccia and Garibaldi (2005) p.g. 671.

[9]The value of $g_{i,t}$ lies in $[-2, 2]$ interval by construction. A value of -2 corresponds to an exit, a value of +2 to an entry.

[10]When bank i is not involved in a merger then $\Delta\tilde{l}_{i,t} = \Delta l_{i,t}$ as can be seen from (2.1).

normalized by total loans outstanding at $t-1$. From (2.4) and (2.5) it is clear that aggregate measures of creation and destruction are weighted averages of credit growth in the set of expanding and contracting banks.

Charter switches and mergers between national and state banks in our sample imply that in some quarters in the cross-section of banks there is a bank(s) for which $g_{i,t}$ is either undefined or it is computed ignoring the merger correction. Including an unadjusted estimate of $g_{i,t}$ in the cross-sectional mean will introduce an upward bias in the measure of credit creation and destruction. This is unappealing because, at least in principle, all reallocation may be driven by this spurious form of credit creation and destruction. We therefore choose to redefine (2.4) and (2.5) to be the weighted sum of credit growth in the set of banks for which this is observed.

More formally, let I_t^+ (I_t^-) be set of banks in quarter t for which the change in credit is positive (negative) and that in that quarter are not involved a state to national merger or charter switch. As well, denote with I_t the union of these two sets[11]. Thus we compute credit creation and destruction as:

$$\tilde{POS}_t = \sum_{i \in I_t^+} \tilde{w}_{i,t} g_{i,t} \tag{2.8}$$

$$\tilde{NEG}_t = \sum_{i \in I_t^-} \tilde{w}_{i,t} |g_{i,t}| \tag{2.9}$$

where

$$\tilde{w}_{i,t} = \frac{0.5(l_{i,t-1} + l_{i,t})}{\sum_{i \in I_t} l_{i,t-1}} \tag{2.10}$$

The extent and nature of the bias that this procedure induces in the measurement of aggregate creation and destruction will depend on the mechanism that drives state banks to merge with national banks or to switch charter. When this transition mechanism is exoge-

[11] By definition the set of banks I_t is a subset of the total number of banks in our sample for each quarter. When in a quarter there are no state to national mergers or charter switches the sets coincide.

nous to contemporaneous credit growth our measure is unbiased. We explore this issue in more detail in section 2.4.4.

Finally, credit reallocation (SUM) is defined as the total amount of credit that is created and destroyed. Excess credit reallocation (EXC) is the part of credit reallocation that is in excess of the net change in credit (NET) expressed in absolute value.

$$SUM_t = \tilde{POS}_t + \tilde{NEG}_t \tag{2.11}$$

$$NET_t = \tilde{POS}_t - \tilde{NEG}_t \tag{2.12}$$

$$EXC_t = SUM_t - |NET_t| \tag{2.13}$$

2.4 Gross Credit Flows

In this section we first characterize the empirical properties of gross credit creation, destruction and excess reallocation in our sample. Second, we study the empirical properties of gross credit flows across secured and unsecured credit. Third, we revisit the impact of compositional effects in driving heterogeneous behavior by isolating the influence of regional shocks. Finally, we perform two robustness checks.

2.4.1 Creation, Destruction and Reallocation

We structure our empirical analysis around three stylized facts about gross credit flows documented in the literature. First we examine mean levels of credit reallocation. Second, we illustrate the volatility properties of the flows and, third, their cyclical characteristics. From a qualitative perspective our results are mostly in line with the literature. To facilitate the quantitative assessment of our results, we compare them to the three main existing contributions to this empirical literature (Dell'Arriccia and Garibaldi (2005), Craig and Haubrich (2013), and Contessi and Francis (2010)). Note, however, that these papers differ from ours in at least three dimensions. First, they document gross credit flows for the entire United States. We instead document gross credit flows for state chartered banks in the state of New York. Second, they study the period (or subperiods) between 1960 and

2008. Our sample covers the period between 1912 and 1932. Third, the literature documents gross credit flows using balance sheet information obtained from the Call reports. We instead use a dataset based on the annual report of the superintendent of the banking department of the state of New York.

In line with the literature, we find that credit creation and destruction coexist across the business cycle. This can be seen in figure 2.2, where we plot credit creation, destruction and net credit growth over our sample period. Both creation and destruction remain positive and substantially above zero throughout. Across the sample, mean net credit growth of 1.09% is the result of the simultaneous expansion of credit in the order of 4.74% and credit destruction of about 3.64% (Table 2.2 column (1)). Excess reallocation of credit is in the order of 4.71%. This implies that about 5% of loans outstanding in the previous period are reallocated across banks in addition to any change in net credit. To put our results into context we report in table 2.2 the mean and standard deviation of gross credit flows in our sample (column (1)) and in the literature (columns (2)-(4)). We find levels of credit creation and destruction which are substantially higher. In particular, both credit creation and destruction in our sample are about 2% higher than those measured by any of the other studies. The exception is credit creation measured by Contessi and Francis (column (4)). However, this result is driven by the substantially higher level of mean net credit growth in their sample relative to all other samples (3.08% vs. 1-2%). Their sample includes the credit boom of the 2000s but not bust phase since the sample ends in early 2008. Our sample on the other hand includes both the 1920s boom and the bust in the late 1920s and early 1930s. Looking at excess reallocation we confirm our result that our sample exhibits substantially higher levels of heterogeneity in bank behavior. In fact, mean excess credit reallocation is 4.71% in our sample, while the literature estimates its range to be between 2.64% - 2.69%.

The fact that we find higher levels of creation, destruction and excess reallocation, however, is not surprising. In fact, we have argued that the benefit of studying this sample period stems from the unit banking structure of the sector. This implies that in our sample we are, in practice, able to measure credit reallocation at the branch level while the Call report data, which underlies the results documented in the literature, only allows the researcher to characterize credit reallocation at the bank holding company (BHC) level. Since BHC level data aggregates balance sheets across multiple branches, the smallest unit of observation in our sample is substantially smaller than that in literature. We believe that this allows us to measure heterogeneity in behavior more precisely and that this drives our results in table 2.2. Additionally, we believe that another advantage of this sample relative to that studied in the literature is the lower level of concentration in the banking system. In fact, conditional on the unit of observation, heterogeneity is better measured when the banking sector is less concentrated. The concentration of the banking sector in this period is in fact far lower (see table 2.1) that that studied by the literature. The Herfindal index of banking concentration in our sample fluctuates around 0.04-0.05, while it fluctuates in the range 0.5-0.9 between 1979 and 1999[12]. This lower concentration rate is in part due to the unit banking nature of the system. Finally, we note that it is possible that the results we document are driven by the bias introduced into our sample by national to state charter switches and mergers. We discuss this issue in section 2.4.4.

We now turn to the volatility properties of the series. Two facts are noteworthy. First, absolute levels of volatility for all series are far higher in our sample than in the literature. Second, we find that credit destruction is more volatile than credit creation, however, the difference between the two series is not as strong as that documented in the literature. In fact, in our sample the coefficients of variation for creation and destruction are respectively 0.58 and 0.71. In the literature, the coefficients of variation for credit creation (destruction) range from .3 to .33 (.41 to .53)[13]. Thus, in the literature, destruction is between 30% to 50% more volatile than creation but only 20% more volatile in our sample. This asymmetry in volatilities is a key feature of credit friction models[14]. Our results suggest that the

[12]See Dell'Arriccia and Garibaldi (2005), p.g. 668, table 1.

[13]Coefficients of variation for credit destruction are .42, .53 and .41 respectively for DAG, CH and CF. Coefficients of variation for credit creation are .32, .33 and .3 respectively for DAG, CH and CF. Computations are based on the results reported in table 2.2.

[14]The asymmetry arises because after a positive aggregate shock credit expansion cannot react immediately since it takes time to screen projects. In response to a negative shock, however, the termination of a credit relationship is not subject to this constraint.

difference between the two margins is not as marked in this period. This may be due to the existence of closer informational ties between banks and clients in the unit banking system that makes creation less subject to moral hazard. When we exclude from our analysis New York City banks (see section 2.4.4) the volatility of creation and destruction are almost identical.

There are two possible explanations for this difference in the absolute volatility we document. First underlying aggregate activity might be more volatile due the difference in the fundamental volatility of aggregate shocks or the policy environment (e.g. lack of countercyclical policy). In line with the results of Romer (1991), we find strong evidence of this in the data. The standard deviation of industrial production q-on-q growth between 1919:Q2 and 1932:Q4 is 6.44%. In the sample period studied by Dell'Arriccia and Garibaldi it is 1.4%. Thus our sample exhibits aggregate business cycle volatility that is roughly 4 times larger[15]. This difference in scale matches exactly the difference in magnitude of the standard deviation of net credit growth in our sample (4.84%) relative to the literature (1.46%, 1.52%, 1.66%). Second, we argue that the lack of branching may also explain some of the time-series variation we document. In fact, in the presence of idiosyncratic credit demand and supply shocks, branching allows bank holding companies to diversify these shocks. On the other hand, when bank are organized as individual and independent units, there is potential mismatch between the supply and demand of credit, which may induce additional variation into the aggregate (Gabaix (2011)).

We now proceed to characterize the cyclical properties of credit flows. Implementation of this in a way that is comparable to the literature, however, is not trivial. In fact, in this historical period (1912-1932) nominal GDP growth[16], the measure of cyclicality used in the literature, is not available on a quarterly basis. GDP figures are only available at the annual frequency for our sample period. Another natural proxy of the national business cycle is

[15]This difference in aggregate volatility persists even when compared to the Craig and Haubrich and Contessi and Francis samples. The standard deviation of industrial production is respectively 1.6% and 1.64% in their samples.

[16]This literature employs nominal GDP instead of real GDP since credit creation and destruction is measured in terms of nominal loan values.

industrial production which is available monthly starting January 1919. However, this is a national measure of the business cycle. Our measures of gross credit growth on the other hand are defined at the state level. We therefore also proxy state level business cycle variation with nominal retail sales for the New York Federal Reserve district. Though the data does not strictly refer to the New York State, the state represented no less than two-thirds of the population in the Federal Reserve district in this period. Data on nominal retail sales is obtained from the Federal Reserve bulletin and is available starting from 1919:Q1. This data series is seasonally adjusted using the Census X-11 procedure as implemented by EViews.

In figure 2.3 we plot year on year growth rates for log industrial production, log retail sales and net credit. We have superimposed the NBER recessions in grey. Both cyclical measures are available only from 1919:Q1, thus the first two recessions in the sample are excluded. Of the national recessions in this period, only in the 1920-1921, 1923-1924 and 1929 recessions do retail sales slow or fall. During the 1926-1927 national recession, however, retail sales maintain a roughly constant growth rate. The time-series pattern in retail sales is also in line with net credit growth. In fact, there are slowdowns in credit growth in the 1920-1921, 1923-1924[17] and 1929 recessions, but not in 1926-1927 recession. For all results on the cyclicality of flows we report results for both the industrial production series as well as retail sales.

In table 2.3 we report the contemporaneous correlation between industrial production and retail sales and our various measures of gross credit flows. In line with the literature, we find that gross creation is pro-cyclical while destruction is countercyclical. On the other hand, we find that excess reallocation is strongly pro-cyclical instead of countercyclical. The pro-cyclicality of excess reallocation seems to be driven by the strong pro-cyclical nature of credit creation. This is in sharp contrast with the results of Dell'Arriccia and Garibaldi (2005), who find that the countercyclicality of excess credit reallocation is driven

[17]This is not clearly visible in the figure using year on year changes. Looking at quarterly changes, however, there is a sharp but brief contraction of credit in the period in question.

by the countercyclical nature of credit destruction.

The latter authors suggest that the countercyclicality of excess credit reallocation is a result of the cleansing effect of recessions. Excess reallocation of credit occurs as banks actively reduce lending to unsuccessful firms and increase lending to new projects to reflect the changed prospects of the firms in their portfolio. Through the lens of their model (Dell'Arriccia and Garibaldi (1998))[18], recessions can therefore be characterized as periods in which the loan portfolio is skewed towards loans with high separation probability and high returns. This reallocation activity, however, raises the riskiness of the loan portfolio since credit restructuring is subject to transaction costs and potential credit losses. A key difference between the historical period studied by Dell'Arriccia and Garibaldi and ours is that deposits were not federally insured prior to 1933[19]. Due to the lack of deposit insurance, during recessions, depositors withdrew liquidity from banks in response to the higher default risk of the banking institutions (Pennacchi (2006)). It is thus reasonable to assume that banks respond to the threat of deposit outflows by choosing not to increase the riskiness of their balance sheet by turning over their portfolio during the recession. In fact, in the recession, the firm is already subject to higher default probability due to the aggregate state. It may thus choose to optimally liquidate unsuccessful but viable investments during periods in which the bank's default probability is low. In essence, pro-cyclical excess reallocation may occur because banks delay the portfolio reallocation that is induced by the recession because they spread credit risks across the cycle (e.g. default risk in the recession, reallocation risk in the expansion). Another related possibility is that credit restructuring is more costly in the recession (e.g. if firms are temporarily illiquid) than in the expansion. In this setting, banks may prefer to liquidate low profitability investments with delay and forgo the additional profits of funding better projects to avoid the increased risk and losses from credit restructuring in the recession.

[18]In such a model banks must choose whether to invest in risky projects or in risk less money market instruments. These projects are ordered by their risk adjusted return so that project profitability can be ranked. As a consequence, the set of risky projects that gets funded increases when the return to money market instruments falls. Excess reallocation of credit occurs through two channels. First, for a given project type, existing projects fail and new ones are funded in its place. Second, the set of funded projects varies in function of the interest rate and this affects excess reallocation since it depends on the total loans outstanding.

[19]The FDIC was instituted with the Banking act of 1933 (Preston (1933)).

2.4.2 Gross credit flows and collateral

A distinctive feature of our data is the category breakdown of total loans outstanding. In fact, loan categories in our dataset are defined in function of the collateral type that secures the loan rather than the purpose of the loan. This is a particularly interesting feature of the data when studying the properties of gross credit flows. This is because collateral alters the informational asymmetries that exist between borrowers and lenders (Bester (1987)). Additionally, it may as well affect the cost and speed with which credit relationships may be terminated (e.g. if the lender can seize the collateral). This difference in the nature of the extension of new loans and the cancellation of non-performing assets is what motivates the literature on gross credit flows. It stands to reason that across collateral type, the properties of creation, destruction and excess reallocation will also vary since collateral affects the parameters that govern these processes. In this section, therefore, we illustrate how the properties of gross credit flows vary in function of the underlying collateral and try to interpret these results in light of a benchmark credit frictions model like that of Dell'Arriccia and Garibaldi (1998).

In our balance sheet data there are four loan categories: loans secured by bonds and real estate, loans secured by mortgages, loans secured by other collateral and unsecured loans. We aggregate into one category the first two types on the basis that the underlying collateral is similar. We refer to this category as loans secured by bonds and real estate. We treat as separate categories loans secured by bonds and real estate and loans secured by other (stocks in large part) due the different volatilities and cyclical properties of the underlying collateral. Given these collateral categories, we would expect the ease with which a credit relationship is formed, and hence the matching rate, to decrease as we moved from loans secured by bonds and real estate, to loans secured by stocks to unsecured loans. This is because we expect that the ease with which credit relationships are formed to be increasing in the quality of collateral. Unsecured loans have no collateral and therefore have the lowest matching rate. Loans secured by stocks are likely to have a lower matching parameter than equivalent loans secured by bonds given the volatility and perhaps pro-cyclicality of stock values.

In table 2.4 we report mean and standard deviation of creation, destruction and excess reallocation for the three loan categories. Three main facts emerge. First, mean creation, destruction and excess reallocation across the three loan categories are higher than their equivalent measures for total loans in table 2.2. This is a symptom of the fact that within banks there is reallocation of credit across the three categories. That is, credit secured by real estate and bonds may be expanding exactly when unsecured credit is contracting[20]. Further, such reallocation is likely to be occurring in function of the business cycle as we document below. Second, we find that the relative volatility of credit creation and destruction varies across loan categories. In particular, we find that for both loans secured by bonds and real estate as well as loans secured by stock, credit creation is relatively more volatile than credit destruction. The coefficients of variation for creation (destruction) are respectively .61 and .64 (.45 and .48). In the unsecured segment we instead find that destruction is relatively more volatile (0.59 against .51) as documented by Dell'Arriccia and Garibaldi (2005). Third, creation, destruction and excess reallocation for loans secured by stocks and unsecured loans is higher than that of credit secured by bonds and real estate.

In table 2.5 we report the contemporaneous correlation between our cyclical measures, retail sales and industrial production, and gross credit flows. A pattern that emerges from the data is that creation and destruction of credit have different cyclical properties across loan categories, though these are statistically significant only for industrial production. In particular, we find that credit creation is pro-cyclical for loans secured by bonds and real estate, but it is almost acyclical for unsecured loans. On the destruction side, we find that for loans secured by bonds and real state the destruction rate is almost acyclical while it is countercyclical for both loans secured by stock and unsecured loans.

[20] To see this consider a bank which is expanding credit in the secured segment and contracting it by an equal amount in the unsecured segment. At the loan category level, this bank contributes to both creation, in secured loans, and destruction, in unsecured loans. However, across categories the bank neither expands nor contracts. When a number of banks are engaging in such activities, the mean creation and destruction by loan category will result to be higher than that of the aggregate.

The key issue is to what extent the results outlined in tables 2.4 and 2.5 are consistent with our priors on the matching parameter. We start by noting that with regards to mean levels of creation, destruction and excess reallocation the model does not make clear cut predictions. In fact, while the increase in the matching rate directly increases mean levels of creation, it also has two indirect effects. In fact, the increased ease with which banks encounter projects implies that (i) banks only fund the most profitable projects (ii) when the number of total projects available in the economy is fixed, the number of unmatched projects available for a match falls. Thus, the effect on creation is undetermined and mean levels of the variables cannot be used to discriminate characteristics of the model. Further, since in the steady state of the model of Dell'Arriccia and Garibaldi creation and destruction are equal, the effect of the matching parameter on excess credit reallocation is also undetermined.

Turning to the properties of credit creation, which the matching parameter directly influences, two results are consistent with our priors. First, the coefficient of variation on creation is higher for loans secured by bonds and real estate than for unsecured loans. In fact, if these loan categories are subject to common aggregate shocks, then this is the expected result since the matching coefficient controls the strength of the response of creation to aggregate shocks. Second, the pro-cyclicality of creation is stronger for loans secured by bonds and real estate than for unsecured loans. Since creation is a frictioned process, the effect of a credit demand shock (e.g. an increase in the number of projects to fund) is distributed over multiple periods as the demand imbalance is only slowly accommodated. The slower the process with which the shock unwinds (e.g. low matching rate), the more the credit friction attenuates the positive correlation between creation and the original shock.

Differences in the matching parameter alone, however, cannot explain a number of other results we find. First, the fact that credit destruction is acyclical for loans secured by bonds and real estate and countercyclical for unsecured loans. Second, the difference in the relative volatility between creation and destruction that exists across loans secured by bonds and real estate and unsecured loans. These findings suggest two characteristics of a model that can rationalize our results. First, the destruction process is also subject to frictions (e.g. administrative bottle-necks, delays in credit recovery). Second, differences in collateral not only affect the ease with which credit relationships are formed but also the ease with which they are destroyed. As in the above discussion, frictions on the destruction side help rationalize both the cyclical properties and the relative volatility of credit destruction. Further, we note that the empirical properties of creation and destruction are such that a high matching rate is associated with a low separation rate (and vice versa). Therefore, any mechanism by which collateral aids the formation of credit relationships but impedes their separation squares off the empirical results described in this paragraph and the previous. A simple way in which this could arise is when loan quality is positively correlated with collateral type (e.g. because collateral signals quality). In creation, collateral reduces informational asymmetry and therefore allows for faster and more cyclical creation of loans. On the destruction side, collateralized loans default mostly for idiosyncratic reasons, hence the lack of cyclicality in destruction. For uncollateralized loans, creation is an acyclical process due to screening, while destruction is countercyclical due to the high sensitivity of these loans to aggregate shocks.

Finally, we note that loans secured by stocks display far higher levels of cyclicality then the other categories. We believe that this result is driven by the fact that, for this loan category, the collateral value is cyclical. The strong pro-cyclicality of creation may be driven by the fact that the screening process is lax when collateral values are high and/or collateral is abundant. As in section 2.4.1, the pro-cyclicality of excess reallocation instead is driven by the desire to liquidate and reallocate the portfolio when collateral values are high to limit potential credit losses. Another interesting finding is that excess credit reallocation is acyclical for both loans secured by bonds and real estate and unsecured loans. This suggests that the matching rate and, based on our above interpretation, the separation parameter do not influence the cyclicality of excess reallocation.

Overall, our results suggest that collateral does influence the probability of establishing a credit relationship. Moreover, they suggest an inverse relationship between the matching parameter and that governing credit destruction. The results also suggest that the cyclicality of collateral drives at in least in part the cyclical properties of gross credit flows. In fact, excess reallocation displays significant cyclical variation only for loan secured by stocks.

2.4.3 Heterogeneous regional shocks

As has been pointed out in the literature, the heterogeneity in gross flows we document in section 2.4.1 may arise due to compositional effects. In particular, Dell'Arriccia and Garibaldi (2005) study the extent to which excess reallocation arises, amongst others, due to heterogeneous regional shocks. A key difficulty in carrying out such analysis with their data, however, is the fact that Call reports possess almost no information about the geographic dispersion of the bank's economic activities. The information they contain is mostly related to the location of the headquarters of the bank. However, when loan decisions are made at the branch level and branches span multiple states, identifying the regional shock with the state in which this bank is headquartered does not cleanly identify regional shocks.

In this dimension our data allows for far cleaner identification of regional shocks. In fact, the banks we analyze are unit banks. In addition to that, banks made loans only within their local market due to informational asymmetries (Bernanke (1983)). This implies that there is a clear correspondence between the bank's location and its economic activity in our data. We therefore collect data on hamlets, towns, villages and cities in the state of New York from the New York State Library to map the location of the banks into their respective county and region within the New York State. We are able to match all but 2 banks in our sample. There are 62 counties in the state of New York and 9 regions. Counties are regional aggregations that are too small to be suitable for our analysis since in many counties we observe no more than a few banks. Due to the limits of the methodology, if all regional aggregations contained only one bank, then all variation in credit reallocation would be attributed to regional shocks. We therefore must strike a balance between the level of aggregation and the heterogeneity in the shocks. Regions are the next level of aggregation and we therefore perform our analysis at this level. Across these 9 regions there is substantial heterogeneity in the number of banks. The New York City region stands out. Alone it represents 35-45% of the banks in our sample. This is not surprising since New York City was a major financial center at the time. On the other hand, even in the region with lowest number of banks (Eastern Adirondacks) there are 15 banks. On average there are 61 banks in each region.

To decompose the proportion of flows that are attributable to heterogeneous shocks at the regional level we borrow the methodology of Davis and Haltiwanger (1992) and Dell'Arriccia and Garibaldi (2005). In particular, we define:

$$\text{Between} = \frac{\sum_j^J |\frac{\sum_t^T NET_{j,t}}{T}|}{\sum_j^J (\frac{\sum_t^T SUM_{j,t}}{T})} \tag{2.14}$$

where j is the region subscript. If within a region there is only credit creation or destruction then it must be that the sum of creation and destruction equals their difference - i.e. $SUM_{j,t} = |NET_{j,t}|$[21]. If this occurs in every region then reallocation is fully driven by regional shocks and the between index takes a value of 1. We find that about 24.7% of aggregate credit reallocation can be explained by heterogeneous regional shocks. This estimate increases to 27.3% when we exclude the New York region from the set of regions J. To put this figure into perspective Dell'Arriccia and Garibaldi (2005) find that only 10% of the excess reallocation is attributable to regional shocks. On the other hand, regional shocks matter far less for bank credit reallocation than for job, sales and firm credit reallocation (Herrera et al. (2011)).

Our decomposition of this variation and that of Dell'Arriccia and Garibaldi are, however, only partially comparable since the studies differ in the time-period under study and in the geographical coverage. In particular, they study heterogeneous shocks at the state level, while we study within state heterogeneous shocks at the region level. To the extent

[21]The quantities are defined in section 2.3

that, within states, there is important heterogeneity then our results are complementary to theirs. On the other hand, if shocks within states are sufficiently homogeneous then our finding that between variation accounts for more of credit reallocation depends on the cleaner identification of the shocks.

Finally, we discuss two potential pitfalls our identification strategy. First, the regions we partition our sample into are administrative divisions. Identification of the regional shocks depends on economic aggregations that are subject to common shocks. Since we cannot investigate the heterogeneity of shocks within regions, we cannot rule out that more of the variation in creation is actually attributable to heterogeneous shocks. Our estimate is therefore downwardly biased in this sense. Second, in regions that have few banks, credit creation and destruction is hard to effectively measure. Since we can measure creation and destruction only across banks, rather than within a bank, regions in which there are fewer banks may appear less heterogeneous than they actually are. In the limit, if all regions had only a single bank then all excess reallocation would be attributed to heterogeneous shocks. In this sense, we may overestimate the importance of regional shocks.

2.4.4 Robustness checks

In this section we investigate to what extent our results may be influenced by banks switching across state and national charter status. We do this in two ways. First we investigate to what extent charter switches are correlated with credit growth. Second, we show how our results are affected when we exclude New York City banks from our sample.

Systematic charter switches

As discussed in section 2.3, we are able to characterize gross credit creation and destruction for New York State only under two key assumptions. First, that state and national banks do not differ systematically in their heterogeneity. Second, that the growth rate of credit for banks switching across charter status is not systematically different from that of banks that maintain a state charter. We cannot perform a direct test of these assumptions, however, we may investigate the association between lagged quarterly (or yearly) growth in credit and charter status switches[22].

In particular, we consider two sources of potential spurious credit growth. First, we consider spurious credit destruction that arises due to state banks being acquired by national banks or state banks moving to a national charter. Second, we consider spurious credit creation that occurs when state banks acquire national banks. For each of these types of events, we check whether lagged quarterly (or yearly) credit growth is a predictor of charter change with a panel probit regression. In the first case, the switchers are defined as state banks moving to a national charter and the stayers are all banks that maintain a state charter. In the second case, switchers are defined as state banks acquiring a national bank, stayers are defined as before. We study these events separately as presumably the selection mechanism operates in diametrically opposite ways. Banks exiting the sample are likely to be experiencing lower levels of loan growth than banks which are actively expanding by acquiring other banks. Finally, note that the two switchers definitions are mutually exclusive but not collectively exhaustive as neither includes national banks switching to a state charter. These cannot be included into the second group as we do not observe credit growth prior to merger. We only observe credit growth for these banks in the periods successive to the merger period.

In both regressions the baseline control set includes the log of outstanding loans in the period previous to the merger and a dummy for the bank being incorporated in any of the five boroughs of New York City. The size of the loan book is used to control for a possible size effect in switching. We include a dummy for banks incorporated in New York City since two-thirds of all the mergers, both state to state and state to national, in our sample occur in the city. We consider two different RHS measure of lagged credit growth: quarterly credit growth - models (1) - (3) - and yearly credit growth - models (4) - (6). Standard errors are computed with bank level clustering. Models (2),(3),(5) and (6) include time

[22]This is not a direct test. In fact, it is always possible that in the merger period banks moving across charters behave systematically differently from other banks, even if lagged/trend credit growth was not systematically different. Nonetheless, since loan growth is a persistent process and mergers are often planned prior to the merger period we think this exercise can be informative of the relationship between charter switches and credit growth.

fixed effects to account for common shocks. The inclusion of fixed effects in panel probit analysis is complicated by an incidental parameters problem (Chamberlain (1984)). In fact, the number of parameters that must be estimated in this non linear model increases with the sample size rendering the MLE estimates inconsistent unless both panel dimensions go to infinity. In our context, however, this problem is attenuated as we estimate time fixed effects in a panel with small T and large N so that the number of incidental parameters remains contained relative to the sample size[23] .

In table 2.6 we report the estimated coefficients for the case of spurious exit. The baseline regression coefficients are reported in columns (1) and (4). We find that there is a positive size effect, and a strong New York City effect. Thus, larger banks switch with higher probability to a national charter and this is especially true in New York City. There is a borderline significant and negative association between lagged quarterly credit growth and charter switches. This suggests that banks exiting the sample are shrinking faster than the average. Note that this would induce a downward bias in our credit destruction measure if the pattern also holds for the merger quarter. The significance of this negative coefficient on credit growth, however, may be driven by differences in average credit growth across time rather than across banks. In fact, a significant proportion of the state to national bank switches occur during the Great Depression. In this period loan growth is significantly more negative than the average. We therefore include time fixed effects to resolve this issue. Once we include them in our regression, the statistical significance of this coefficient goes away, though the magnitude remains roughly constant. We point out that the inclusion of time fixed effects drastically reduces the number of observations. This is because in the vast majority of quarters, there are no state to national charter transitions and thus the time fixed alone perfectly predicts the outcome. Finally, in models (3) and (6) we re-estimate this model using only banks in New York City[24]. We do this because it

[23]In practice, despite the fact that we have 85 quarters of data, there will be far less time dummies to estimate. In fact, we only require estimation of time dummies for quarters in which there are state to national charter switches or mergers. All other quarters are dropped since the time fixed effect perfectly predicts the outcome (i.e. no charter switches). Since the panel is roughly balanced, the proportion of incidental parameters to the sample size remains constant.

[24]For obvious reasons the New York City dummy is dropped in this specification.

is possible that switching is associated with loan growth in New York City but not in the rest of the state where state charters were far more common than national charters. Again we find no evidence of a statistically significant relationship between lagged credit growth and charter switches.

In table 2.7 we report the estimated coefficients for the case of spurious entry. The results are similar to the ones discussed above. We find that state banks acquiring national banks are typically bigger banks. As well, we find no statistically significant relationship between lagged credit growth and this type of merger once we control for time fixed effects (columns (2) and (4)). The sign of the coefficient is positive suggesting that our procedure systematically excludes from our credit creation measure banks which are expanding credit faster in the quarter previous to the merger than the average state bank. Thus we may underestimate credit creation when we exclude these observations. If these patterns also hold in the merger quarter, this suggests that we may underestimate both creation and destruction and in consequence excess reallocation. Thus our finding that average excess reallocation is higher than that measured in the literature does not seem to be driven by our treatment of state to national charter switches. Finally, we could include town fixed effects in our regressions to control for heterogeneity in transition rates across towns. It turns out that the only important heterogeneity in this sense is well captured by our New York City dummy.

Gross credit flows in New York State

As discussed in section 2.3, the major drawback of our sample is the lack of data on nationally chartered banks. This problem complicates our measurement of aggregate gross credit flows since we are unable to characterize the behavior of banks merging or moving across charter status in the cross-section of credit growth. An important characteristic of our sample is that the incidence of this type of mergers or charter status changes is very low outside of New York City. In fact, there are a total of 156 mergers or charter changes in our sample. However, only about a fifth of these events occur outside New York City (34). Further, of these 34 instances 26 are mergers between state chartered banks or trusts and therefore pose no measurement difficulty. Of the remaining 8 events, we have two instances in which nationally chartered banks become state chartered banks and thus enter our sample. We have one instance in which a state bank absorbs a national bank and maintains a state charter[25]. In both cases these events generate spurious credit creation. We have five instances in which state banks merge with a national bank and exit the sample[26]. These events generate spurious credit destruction. In addition, these events are concentrated around a particular set of dates and therefore the vast majority of the time-series is actually not affected by these particular merger events.

We therefore study creation, destruction and excess reallocation in the New York State subsample to investigate to what extent the higher levels of excess credit reallocation and its pro-cyclicality we documented are driven by our inability to measure credit growth for state to national charter transitions. We do this in two steps. First, we show the extent to which total credit creation and destruction in the New York state sample are driven by spurious credit creation and destruction. We then proceed to verify that the results in section 2.4.1 remain true also for this subsample.

In figure 2.4 we plot in red the level of total credit creation (top panel) and destruction (bottom panel). Differently from the previous sections, we include in total credit creation both true credit creation (as before) as well as spurious credit creation. We do this to highlight the instances in which spurious credit creation affects our measurement of gross credit flows. In both panels, in blue we plot two subcomponents of total credit creation (destruction): spurious creation (destruction) and creation (destruction) due to entry (exit). In dark blue we plot the part of credit creation (destruction) that occurs because of regular entry (exit). In light blue we plot instead the creation (destruction) of credit that occurs because of spurious entry, charter switches and exit. The figure illustrates two facts: first, between the creation and destruction series only three observations are contaminated

[25]Mount Vernon Trust Company merges with American Bank and Trust Company of Mount Vernon

[26]In particular we have that Mechanics Bank of Groton merges with First National Bank of Groton, Bingham State Bank merges with Genesee River National Bank and Trust Co, Livonia State Bank switches charter, Yonkers Trust Company merges with First National Bank of Yonkers, Livingston County Trust Company becomes a national bank.

by spurious credit creation and destruction. That is, 82 of the 85 quarters in which we characterize creation, destruction and excess reallocation are independent of how we treat state to national (or vice versa) charter switches. There are a total of 8 instances in which we observe charter switches in this subsample, however, these concentrate around only 3 dates. Second, the overwhelming majority of credit creation and destruction occurs across banks that remain active and keep their charter status. Note as well that if we had misclassified entry or exit of national banks in our merger file these would show up in the entry and exit series. We find no evidence of any misclassification in this sense.

In table 2.8 we report the mean, standard deviation and contemporaneous correlation with IP and nominal retail sales of gross credit creation, destruction and excess reallocation for New York State. We include in the table the results for the full sample from tables 2.2 and 2.3 for comparability. We find that our basic findings are confirmed in this sample though the magnitudes are smaller. In particular, excess reallocation falls to 3.46% from 4.71%. This is still higher than that in the literature (2.6%), but not substantially larger. We also find that the pro-cyclical nature of excess credit reallocation is confirmed thought it is not statistically significant when using industrial production as the cyclical measure in the subsample. Overall, we conclude that the results we have presented throughout this paper do not seem to be substantially influenced by our treatment of national to state charter switches and mergers.

2.5 Conclusion

In this paper we have studied gross credit flows for the state of New York using balance sheet data collected from the annual report of the superintendent of the banking department of this state. We focus our analysis on this historical period, due to the unit banking nature of the financial system. The lack of branching allows us to measure heterogeneity in credit expansion and contraction across banks to a finer degree than has been previously possible. Consistent with this premise, we find higher levels of gross credit creation, destruction and excess reallocation than those documented in the literature. In particular, we find that mean excess reallocation in our sample is in the order of 4.71%, almost twice the estimate of the previous literature. This result is subject to an important caveat. Since we do not observe balance sheets for nationally chartered banks we are unable to characterize credit growth in the merger quarter for the set of banks that switches charter. This may induce bias into our estimates of gross credit flows. To the extent that it is possible given our data, we show that the likely bias in our sample actually attenuates our results rather than magnifies them. Further, when we limit our sample to a subsample in which this problem is very limited we still find higher levels of excess credit reallocation than in the literature, albeit slightly lower ones than in the full sample.

Finally, we also documented the behavior of gross bank credit flows by collateral type. We find that collateralized loans are pro-cyclical in the creation margin, but acyclical in the destruction margin. On the other hand, uncollateralized loans are acyclical in creation but countercyclical in destruction. As well, we find that creation is more volatile than destruction for collateralized loans, while the opposite is true for uncollateralized loans. These results suggest that collateral reduces the informational asymmetry between borrowers and lenders making the creation margin more volatile and responsive to business cycle fluctuations than it would otherwise be. Further, they suggest an inverse relationship between the matching rate and the separation rate and that these two parameters do not influence the cyclicality of excess reallocation.

2.6 Appendix

Figures

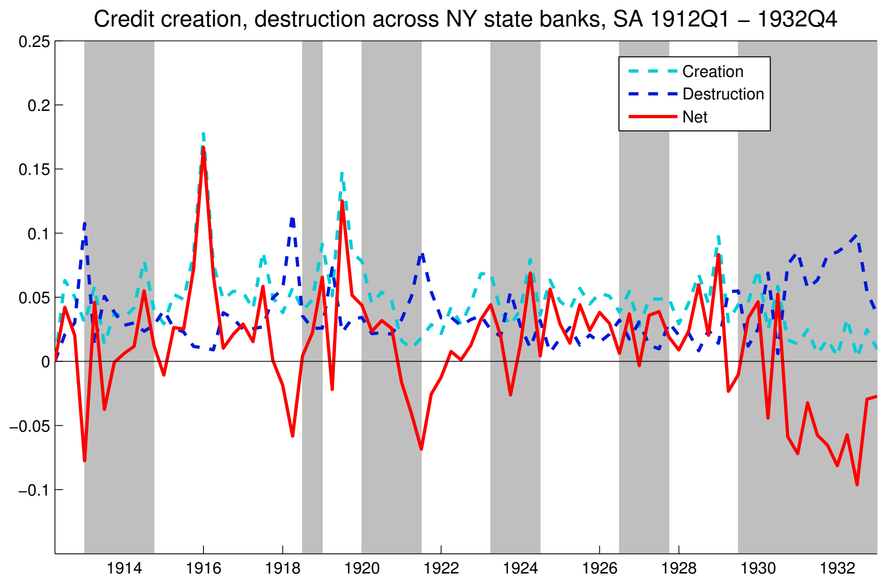
Figure 2.1: Assets and Liabilities Reported

Assets	Liabilities
Specie	Capital
Other currency authorized by the United States government	Surplus, including all undivided profits
Cash Items	Preferred deposits, viz:
Due from NY Federal Reserve Bank, less offsets	Due New York State savings banks
Due from other approved reserve depositories, less offsets	Due New York State savings and loan associations, credit unions, and land bank
Due from other banks, bankers, and trust companies	Deposits by the State of New York
Stocks and bond investments	Other deposits secured by the pledge of assets
Loans and discounts secured by bond, mortgage, deed, or other real estate collateral	Deposits otherwise preferred
Loans and discounts secured by other collateral	Due depositors, not preferred
Loans, discounts, and bills purchased but not secured by collateral	Due trust companies, banks, and bankers
Own acceptances purchased	Bills payable
Overdrafts	Rediscounts
Bonds and mortgages owned	Acceptances of drafts payable at a future date or authorized by commercial lines of credit
Real estate	Other liabilities
Customers' liability on acceptances (per contra, see liabilities)	
Other Assets	

Source: New York State Banking Department (1929).

Figure 2.2: Net credit growth and excess reallocation

The top panel reports the time-series of credit creation, destruction and net credit growth. The bottom panel reports net credit growth and excess credit reallocation. Shaded in grey are the NBER national recession dates.



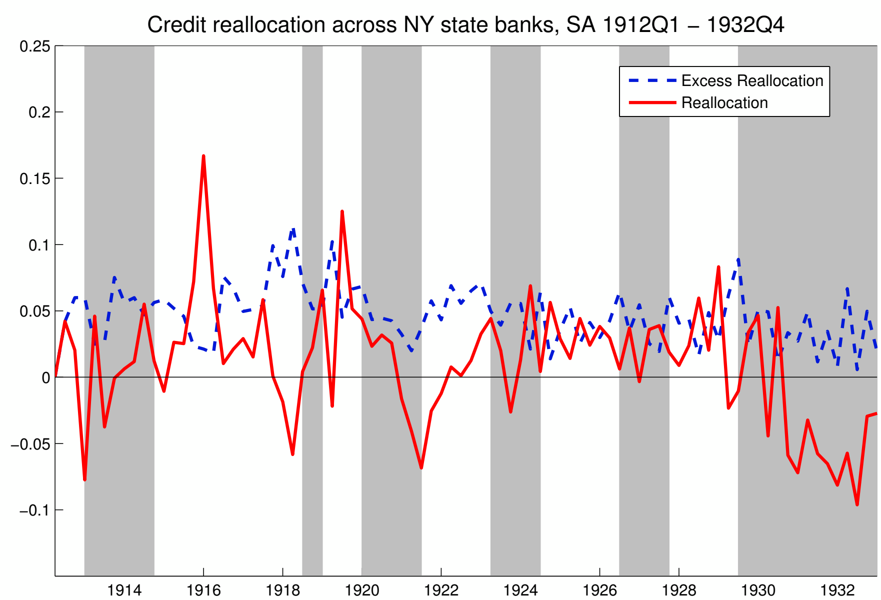


Figure 2.3: Business Cycle proxies: National and State level

Industrial production growth, retail sales and net credit growth are reported as y-on-y growth rates to facilitate comparison. Industrial production is obtained from FRED (mnemonic INDPRO) and is a national figure. Nominal Retail Sales are obtained from the Federal Reserve bulletin and are specific to the New York Federal Reserve district. Shaded in grey are the NBER national recession dates.

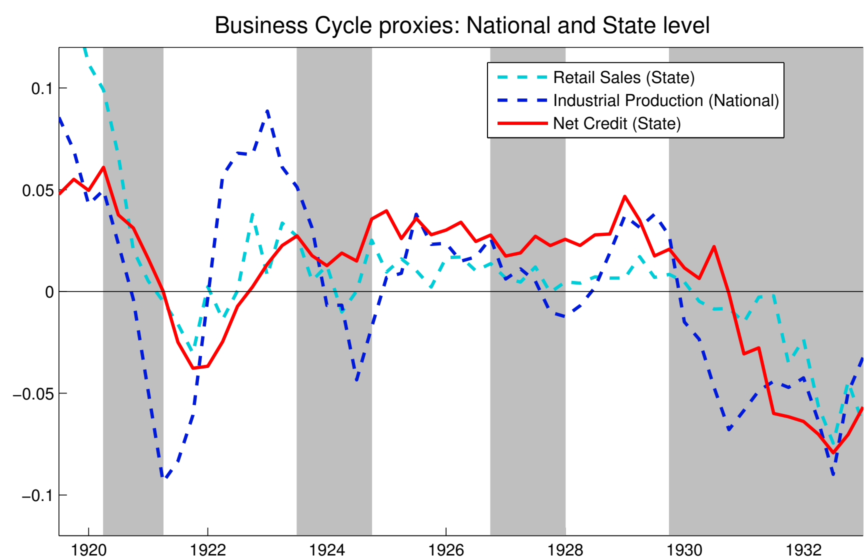
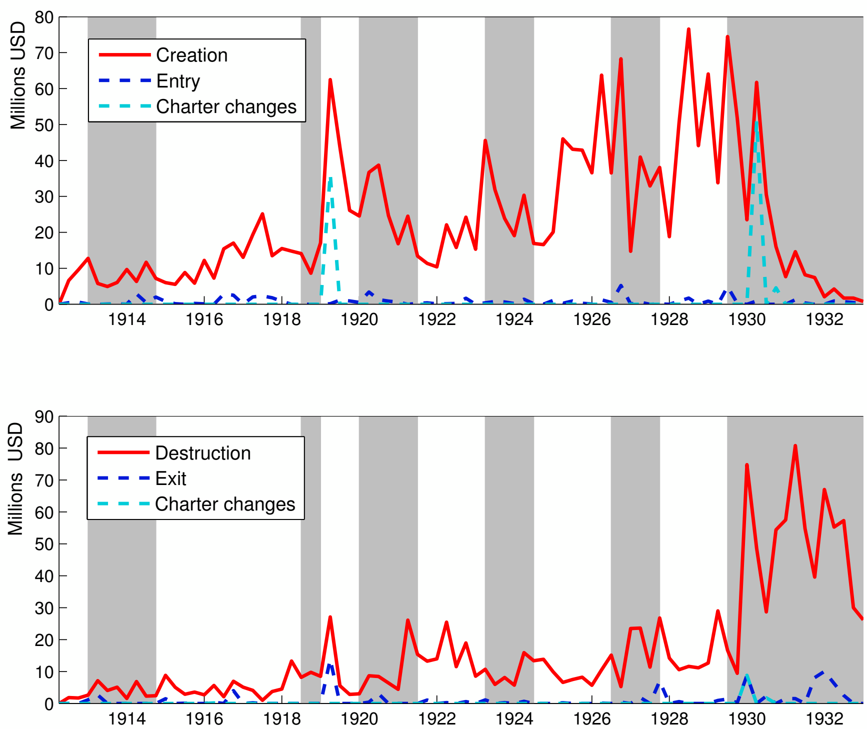


Figure 2.4: Creation and Destruction in NY state

Total credit creation (top panel) and destruction (bottom panel) for the state of New York (in red). In dark blue we plot the creation and destruction subcomponents that arise due to entry and exit of banks. In light blue we plot the amount of spurious credit creation and destruction. Shaded in grey are the NBER national recession dates.



Tables

Table 2.1: Evolution of the State Chartered Banks and Trusts in NY State, 1912-1932

Sample statistics are reported for the 4th quarter of each year. Num. Banks is the number of banks with non-zero assets in the fourth quarter of the year. Agg. Loans is the value of total loans across state banks relative to total loans across state banks in 1912.

Year	Num. Banks	Agg. Loans	Herfindal	Gini
1912	280	100.00	0.03	0.782
1913	277	98.19	0.03	0.787
1914	287	105.60	0.03	0.805
1915	290	138.66	0.05	0.843
1916	296	156.87	0.05	0.841
1917	314	168.12	0.05	0.839
1918	310	181.91	0.06	0.848
1919	324	224.73	0.06	0.845
1920	343	243.43	0.05	0.846
1921	340	203.64	0.04	0.836
1922	344	213.12	0.04	0.843
1923	358	239.60	0.03	0.845
1924	373	278.04	0.04	0.86
1925	378	309.57	0.03	0.851
1926	398	339.93	0.03	0.845
1927	392	381.27	0.04	0.853
1928	393	450.36	0.04	0.866
1929	391	574.56	0.05	0.89
1930	376	448.27	0.06	0.888
1931	364	331.77	0.06	0.88
1932	323	253.65	0.05	0.88

Table 2.2: Creation, Destruction and Excess Reallocation

We report the mean and associated standard deviation in parenthesis of the quarterly growth rate of net credit, gross credit creation, destruction and excess reallocation in our sample and the literature - as reported in Dell'Arriccia and Garibaldi (2005), Table 3, p.g. 675, Craig and Haubrich (2013), Table 1, P.g. 408, Contessi and Francis (2010), Table 1, p.g. 435. Note that we measure credit flows only for state chartered banks and trusts within New York State. The literature reports credit flows for the entire United States. All figures are in percentage points.

		Existing literature		
	CP (2015)	DAG (2005)	CH (2013)	CF (2010)
	1912:Q2 - 1932:Q4	1979:Q2 - 1999:Q2	1960:Q1-2004:Q3	1999:Q1 - 2008:Q2
	(1)	(2)	(3)	(4)
NET	1.09	1.76	.86	3.08
	(4.54)	(1.46)	(1.52)	(1.66)
POS	4.74	3.18	2.91	4.40
	(2.77)	(1.02)	(0.97)	(1.43)
NEG	3.64	1.42	2.04	1.32
	(2.46)	(0.62)	(1.09)	(0.54)
EXC	4.71	2.69	-	2.64
	(2.14)	(0.98)	-	(1.07)

Table 2.3: Creation, Destruction and Excess Reallocation: Correlation with the Business Cycle

All flows are seasonally adjusted using the EViews X-11 procedure. GDP, IP and retail sales (RS) and gross credit flows are the cyclical component of the respective variables. This is defined as the deviation of the log level of the variable from its HP filtered trend. The correlations reported for DAG (2005) and CF (2010) are based on contemporaneous correlation as reported in Dell'Arriccia and Garibaldi (2005), Table 6, p.g. 678 and Contessi and Francis (2010), Table 2, p.g. 438. Industrial production is obtained from FRED (mnemonic INDPRO) and is a national aggregate. Nominal Retail Sales are obtained from the Federal Reserve Bulletin and refer to the New York Federal Reserve district. Standard errors are Newey-West standard errors with lag length set to 4 quarters.

			Existing literature	
	CP (2015)		DAG (2005)	CF (2010)
	1919:Q2 - 1932:Q4		1979:Q2 - 1999:Q2	1999:Q1 - 2008:Q2
	$Corr(\bullet, IP)$	$Corr(\bullet, RS)$	$Corr(\bullet, GDP)$	$Corr(\bullet, GDP)$
POS	.64***	.86***	.35	.38
	(.22)	(.06)	.	.
NEG	-.49**	-.44	-.32	-.46
	(.23)	(.32)	.	.
EXC	.71***	.68***	-.29	-.46
	(.12)	(.06)	.	.

Table 2.4: Creation, Destruction and Excess Reallocation by collateral type

The table reports the mean and standard deviation, in parenthesis, of the quarterly growth rate of nominal gross credit creation, destruction and excess reallocation for loans secured by bonds and real estate, loans secured by stocks, and unsecured loans. All figures are in percentage points.

	Loans secured by:		Unsecured Loans
	Bonds and Real Estate	Other/Stocks	
POS	5.86	7.22	7.04
	(3.54)	(4.61)	(3.68)
NEG	4.25	6.15	5.80
	(1.95)	(2.99)	(3.47)
EXC	7.10	8.34	7.96
	(2.44)	(3.56)	(3.25)

Table 2.5: Creation, Destruction and Excess Reallocation by collateral type: Correlation with the Business Cycle

All flows are seasonally adjusted using the EViews X-11 procedure. IP, retail sales (RS) and gross credit flows are the cyclical component of the log level. This is defined as the deviation of the log level of the variable from its HP filtered trend. Industrial production is obtained from FRED (mnemonic INDPRO) and is a national aggregate. Nominal Retail Sales are obtained from the Federal Reserve Bulletin and refer to the New York Federal Reserve district. Standard errors are Newey-West standard errors with lag length set to 4 quarters.

	Loans secured by:				Unsecured Loans	
	Bonds and Real Estate		Other/Stocks			
	$Corr(\bullet, IP)$	$Corr(\bullet, RS)$	$Corr(\bullet, IP)$	$Corr(\bullet, RS)$	$Corr(\bullet, IP)$	$Corr(\bullet, RS)$
POS	.43***	.11	.68***	.83***	.14	.23
	(.18)	(.28)	(.21)	(.07)	(.14)	(.18)
NEG	-.09	.13	-.47**	-.36	-.34**	-.23
	(.23)	(.13)	(.22)	(.28)	(.18)	(.33)
EXC	.27	.25	.54***	.65***	-.03	.19
	(.26)	(.18)	(.21)	(.18)	(.17)	(.25)

Table 2.6: Spurious Credit Destruction: Probit Analysis of exiting banks

Standard errors clustered by bank are reported in parenthesis. The dependent variable is a dummy that take a value of 1 for state banks that merge with national banks and exit the sample (i.e. become nationally chartered) and for state banks which switch to a national charter. These switches generate spurious credit destruction and there are a total of 25 instances in which this occurs.

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta l_{i,t-1}$	-0.354*	-0.457	-0.491	-	-	-
	(0.211)	(0.303)	(0.404)	-	-	-
$\Delta l_{i,t-4}$	-	-	-	-0.664	-0.921	-0.930
	-	-	-	(0.499)	(0.709)	(0.897)
$l_{i,t-1}$	0.0949**	0.0817*	0.120**	0.0923**	0.0780*	0.114**
	(0.0380)	(0.0443)	(0.0536)	(0.0379)	(0.0445)	(0.0530)
New York City	0.490***	0.730***	-	0.503***	0.750***	-
	(0.163)	(0.193)	-	(0.163)	(0.193)	-
Constant	-4.757***	-4.057***	-3.755***	-4.712***	-4.018***	-3.654***
	(0.548)	(0.701)	(0.965)	(0.545)	(0.704)	(0.951)
Time fixed effects	No	Yes	Yes	No	Yes	Yes
Observations	27,203	5,573	991	27,209	5,573	991

Table 2.7: Spurious Credit Creation: Probit Analysis of entering banks

Standard errors clustered by bank are reported in parenthesis. The dependent variable is a dummy that take a value of 1 for state banks acquiring national banks and maintaining a state charter. These switches generate spurious credit creation and there are a total of 20 instances in which this occurs.

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta l_{i,t-1}$	0.388*	0.449	0.0327	-	-	-
	(0.201)	(0.288)	(0.655)	-	-	-
$\Delta l_{i,t-4}$	-	-	-	0.387*	0.580	-1.153
	-	-	-	(0.230)	(0.394)	(0.854)
$l_{i,t-1}$	0.172***	0.174***	0.280***	0.172***	0.179***	0.282***
	(0.0422)	(0.0491)	(0.0776)	(0.0424)	(0.0502)	(0.0778)
New York City	0.190	0.343	-	0.195	0.336	
	(0.179)	(0.214)	-	(0.179)	(0.214)	
Constant	-5.895***	-5.513***	-6.828***	-5.899***	-5.584***	-6.938***
	(0.631)	(0.837)	(1.476)	(0.635)	(0.850)	(1.491)
Time fixed effects	No	Yes	Yes	No	Yes	Yes
Observations	27,178	5,833	918	27,184	5,834	918

Table 2.8: Creation, Destruction and Excess Reallocation in NYS

In the left side of the table, we report the mean values and associated standard deviation in parenthesis of the quarterly growth rate of gross credit creation, destruction and excess reallocation. Mean and standards deviations are in percentage points. In the right side of the table, we report contemporaneous correlations between HP filtered gross credit flows and our measures of business cycle variation, industrial production and retail sales. Standard errors are Newey-West standard errors with lag length set to 4 quarters.

	Mean and variance		$Corr(\bullet, IP)$		$Corr(\bullet, RS)$	
	Full Sample	NY state	Full Sample	NY state	Full Sample	NY state
POS	4.74	4.27	.64***	.57**	.86***	.86***
	(2.77)	(2.51)	(.22)	(.25)	(.06)	(.05)
NEG	3.64	2.34	-.49***	-.42	-.44	-.34
	(2.46)	(1.46)	(.23)	(.26)	(.32)	(.31)
EXC	4.71	3.46	.71***	.38	.68***	.54***
	(2.14)	(1.89)	(.12)	(.23)	(.06)	(.15)

CHAPTER 3

Price dispersion over the business cycle: Evidence from the airline industry - with Kristopher S. Gerardi and Adam H. Shapiro

3.1 Introduction

Economists have long been captivated by the fact that for many homogeneous goods, a distribution of prices exists rather than a single price. Numerous empirical studies (for example, Shepard (1991), Sorensen (2000), Stavins (2001), and Hendel and Nevo (2011)) and theoretical models (for example, Salop and Stiglitz (1982), Burdett and Judd (1983), Holmes (1989), and Aguirre et al. (2010)) have been produced to better understand this phenomenon, but its fundamental causes are still widely debated. This paper adds to the empirical literature on this topic by providing evidence on how price dispersion moves with the business cycle.[1] Understanding how aggregate factors affect price dispersion may ultimately provide economists with a better understanding of firms' pricing decisions.

The airline industry has been the focus of many empirical studies on price dispersion for a few important reasons. First, the airline industry is one in which firms are well known to charge a distribution of prices for the same product. Thus, there exists a large degree of price dispersion in the industry. Second, markets in the airline industry are cleanly delineated by distinct routes, which allows researchers to empirically assess price dispersion through panel-data methods. Finally, high quality data on airline prices and costs at relatively granular levels are publicly available.

In this study, we examine how various measures of price dispersion at the route level in the airline industry are correlated with the business cycle, while controlling for variation

[1] Additional studies on the determinants of price dispersion include Baye et al. (2004), Goldberg and Verboven (2001), Gaggero and Piga (2011), and Orlov (2011).

in price dispersion that is likely due to other factors, such as market structure, fuel, and cost variations. Our main result is that price dispersion moves pro-cyclically in the airline industry. Using a fixed-effects estimation on a panel that spans almost two full business cycles, we find that a rise in the output gap—a measure of the difference between nominal GDP and "potential" GDP as defined by the Congressional Budget Office—of 1 percentage point is associated with a 1.5 percent increase in the interquartile range, on average. A fall in the average city-endpoint unemployment rate of one percentage point causes a 2.3 percent rise in the interquartile range, on average. Our results are robust to a range of different measures of price dispersion, including the Gini coefficient. Previous studies have found it important to differentiate between legacy carriers, also known as "hub-and-spoke" carriers, and low-cost carriers (LCCs), because they behave quite differently along dimensions related to pricing, competition, and network formation. Interestingly, our results indicate that price dispersion is more pro-cyclical for legacy carriers relative to LCCs.

There are a number of potential mechanisms that could cause price dispersion to move pro-cyclically. With the available data on prices, costs, and purchaser demographics, we are unable to single out any one particular mechanism as the sole contributor to this empirical finding. For instance, since we do not observe many of the individual ticket characteristics we cannot rule out the possibility that price dispersion varies due to a change in the tickets consumers purchase. However, we are able to provide theoretical and empirical evidence that favors some explanations over others. In particular, we provide evidence that pro-cyclical price dispersion may be a simple outcome of second-degree price discrimination tactics. We also provide empirical evidence that downplays the importance of stochastic-demand pricing (Eden (1990)). This result corresponds well with the recent findings of Puller et al. (2009) who find that airline price dispersion is driven primarily by second-degree price discrimination tactics, as opposed to stochastic-demand pricing techniques.

This study is related to the growing literature on price dispersion in the airline industry. It is also related to numerous microeconomic studies on pricing strategies and business cycle conditions. For instance, Rotenmberg and Saloner (1986) theorize that during booms firms may be less likely to collude since the benefits of cheating are higher, causing firms to cut prices. There are a number of empirical papers on this topic that document that retail prices tend to fall during periods of peak demand (see Warner and Barsky (1995), MacDonald (2000), Chevalier et al. (2003), and Nevo and Hatzitaskos (2006)). Another set of theories, based on switching costs and brand loyalty show that during booms new customers may enter the market causing demand to become more elastic and firms to lower prices (see Bils (1989), Klemperer (1995), and Stiglitz (1984)). A third theory, put forth by Greenwald et al. (1984) and analyzed by Chevalier and Scharfstein (1996), shows that during recessions cash-strapped firms may forego offering low prices to attract new customers in order to generate a higher cash flow.

Although our empirical analysis is confined to one industry, we believe it likely has implications for other industries as well. If the correlation between airline price dispersion and measures of the business cycle that we document is due in part to price discrimination tactics, then we would expect to find pro-cyclical price dispersion in industries that are characterized by firms with market power and the ability to implement discriminatory pricing strategies such as hotels, stadiums, restaurants, theaters (Leslie (2004)), yellow-page advertising (Busse and Rysman (2005)), cement (Miller and Osborne (2010)) and personal computers (Aizcorbe and Shapiro (2010)).

The paper is structured as follows: Section 3.2 contains a detailed discussion of the data. In Section 3 we perform a fixed-effects, panel estimation of the relationship between price dispersion and various proxies for the business cycle. In Section 4 we provide a discussion of our empirical findings, paying particular attention to two leading theories of price dispersion: price discrimination and stochastic-demand pricing. We conclude in Section 5.

3.2 Data

The empirical analysis focuses on domestic, direct, coach-class airline tickets over the period 1993q1 to 2009q4. The sample is constructed in the same manner as in Gerardi and Shapiro (2009) and includes nine major domestic airlines, often referred to as "legacy carriers,"[2] as well as a number of low-cost carriers[3] (LCCs) and regional carriers. Ticket prices represent 10 percent of all domestic tickets issued by airlines and are obtained from the DB1B database. In addition to ticket prices, the DB1B includes other quarterly itinerary information, such as origin and destination airports, passenger quantities, number of stops (plane changes), and fare class.[4] Tickets less than 20 dollars are believed to be frequent-flyer tickets and are eliminated.

The data is a panel, where an observation is a flight conducted by a specific airline, between an origin and destination airport (route), in a specific time period (year and quarter). For example, an American Airlines direct, coach-class ticket, from Dallas (DFW) to San Francisco (SFO) in the first quarter of 1999 is considered an observation in our data. The direct ticket data include both one-way flights and round-trip flights. The DB1B contains numerous itineraries and fares for the same flight by the same carrier, reflecting the quarterly frequency of the data, as well as the many different fares found within the same fare class, on the same flight, at a given point in time. Thus, the data comprise distributions of prices for carrier-route itineraries.[5] Price dispersion is measured using three separate proxies: the interquartile range, the Gini coefficient, and the 90th and 10th price percentiles estimated separately. The interquartile range and Gini coefficient are advantageous in that they summarize dispersion with one statistic, while the price percentiles have the advantage that they provide more detailed information about the tails of the distribution.

Table 3.1 displays summary statistics of the variables that we include in our regression analysis. The mean Gini coefficient in our entire sample is 0.22, and is 0.25 for legacy carriers and 0.17 for LCCs. The Gini coefficient can be shown to be equal to twice the expected absolute difference between two ticket prices drawn randomly from the population. For

[2]The legacy carriers in our sample include United, US Airways, Delta, American, Alaskan, TWA, Continental, Northwest, and America West.

[3]The list of LCCs, obtained from Ito and Lee (2003), includes Air South, Access Air, AirTran, American Trans Air, Eastwind, Frontier, JetBlue, Kiwi, Morris Air, National, Pro Air, Reno, Southwest, Spirit, Sun Country, ValuJet, Vanguard, and Western Pacific. For a more detailed discussion of LCCs see Goolsbee and Syverson (2008).

[4]There are three different sub-components to the DB1B data set. They are market data, coupon data, and ticket data; and we combine variables from all three. For further reference, see the BTS's website http://www.transtats.bts.gov.

[5]See Appendix B for more details on the construction of the dataset, and Gerardi and Shapiro (2009) for an even more detailed description.

example, the median Gini coefficient for the entire data set is 0.225, which corresponds to an expected fare difference of 45 percent of the mean fare for two randomly selected passengers. The mean interquartile range (IQR) is 92 dollars for our entire sample, and is 112 dollars for legacy carriers and 52 dollars for LCCs. Figure 3.1 plots the passenger-weighted average of the IQR over the sample period, along with the output gap, as measured by the Congressional Budget Office (CBO). The average degree of price dispersion rises with the boom in the late 1990s and then falls with the ensuing recession. Dispersion is flat throughout the mid 2000s, during which time the output gap was roughly zero, and then dispersion falls somewhat with the latest recession in 2008.

We present a few graphical examples of the pricing patterns seen in the data in order to show a more granular detail of the dynamics of price dispersion. Figure 3.2 plots price percentiles of three routes along with a plot of the output gap. The output gap is defined as the log difference between the actual nominal GDP and the CBO's measure of potential output. The top two panels correspond to routes operated by two legacy carriers, American Airlines and Delta Airlines, while the bottom panel consists of a route operated by Southwest Airlines. It is noteworthy that in the legacy carrier panels, the higher price percentiles seem to closely follow the output gap. The top portion of the price distribution rises and falls with the boom in the late 1990s and then begins to gradually fall as aggregate demand deteriorates. In contrast, we do not see the same relationship in the Southwest panel.

3.2.1 Operating Cost

As we are interested in studying variation in price dispersion that cannot be explained by variation in airline operating costs alone, we must include a control for the airlines' marginal cost in our empirical analysis. Airline marginal costs may vary over the business cycle for many reasons. For instance, wages of pilots and flight attendants may rise during booms, as may the price of fuel. We proxy for variations in marginal cost using a measure of the carrier's average variable cost. Numerous studies, such as Caves et al. (1984) and Gillen et al. (1990), have found that the carriers' passenger output displays constant-returns-to-scale in firm size. This finding suggests that average variable cost may be a valid approximation to marginal cost in this context. We exploit the rich cost data available in the BTS P-52 database. Specifically, the BTS defines a measure called the "total aircraft operating cost," which includes fuel, crew wages, maintenance, aircraft leasing, and depreciation. We are also able to decompose this variable into its fuel component and its other components. Due to the large market power of unions in the airline industry, non-fuel costs are particularly rigid relative to fuel costs.

Figure 3.3 plots total aircraft operating cost (including fuel) as a proportion of total seat-miles for four carriers over the sample period. The figure shows that cost per seat-mile is correlated across firms, and has generally increased through the course of the sample period. The large rise and fall in costs in 2008 can be attributed to the spike in oil prices that occurred during that summer. Southwest and JetBlue, the two largest LCCs in our sample, have lower cost levels relative to the two legacy carriers, US Airways and United. This differentiation in cost between legacies and LCCs is ubiquitous across the entire airline industry. Table 3.1 provides summary statistics for our cost measures used in the empirical analysis. Total aircraft operating cost (less fuel) as a proportion of total seat-miles, $COST$, are higher on average for legacy carriers: 3.4 cents per seat-mile for legacy carriers as opposed to 2.8 cents per seat mile for LCCs. However, fuel costs ($FUEL$), measured as price per gallon, are higher for LCCs. Overall, including a proxy for marginal cost in the empirical specification removes any variation in price dispersion induced by variation in tangible costs.

3.3 Estimation

Since the data is a panel of airline-route observations, it is possible to assess the effects of business cycle variation on price dispersion while holding fixed time-invariant, route-specific factors, as well as any route-specific variation in the degree of competition and carrier-specific variation in fuel and other operating costs. We use a fixed-effects panel estimator, which exploits the time-series variation along a specific route in the estimation routine. We use two different approaches to measure the effect of business cycle variation on price dispersion.

The first specification takes the form:

$$\text{DISP}_{ijt} = \theta_0 + \beta * YGAP_t + \gamma_1 * \widehat{HERF}_{jt} + \gamma_2 \ln FUEL_{it} + \gamma_3 \ln COST_{it} + \delta_q + \nu_{ij} + \varepsilon_{ijt} \quad (3.1)$$

where i indexes the carrier, j the route, t the specific time period, and q the quarter. In this specification, the output gap, $YGAP_t$ is used to proxy for the business cycle, as measured by the CBO, and carrier-route fixed effects are represented as ν_{ij}. We include the Herfindahl index, $\widehat{HERF}_{jt}$, to control for variation in market concentration of the route. As this measure is endogenous, we instrument using the same variables as in Borenstein and Rose (1994) and Gerardi and Shapiro (2009). These instruments include the total number of enplaned passengers on the route, a measure of predicted concentration, and a measure of the airline's share of enplaned passengers at both endpoints. These variables are meant to capture exogenous variation in the degree of competition that are not directly correlated with the firm's pricing decision. We control for time-series variation in costs on a specific carrier i with the logarithm of the carrier's average fuel cost per gallon, $\ln FUEL_{it}$, as well as the remaining operating cost per seat-mile, $\ln COST_{it}$, measured by the BTS for a specific carrier. We also include quarter dummies, δ_q, to control for seasonal fluctuations.

The second specification takes the form:

$$\text{DISP}_{ijt} = \theta_0 + \beta * UR_{jt} + \gamma_1 * \widehat{HERF}_{jt} + \gamma_2 \ln FUEL_{it} + \gamma_3 \ln COST_{it} + \delta_q + \nu_{ij} + \varepsilon_{ijt} \quad (3.2)$$

where the average unemployment rate of the two endpoint states on the route obtained from the Bureau of Labor Statistics, UR_{jt}, is used as an alternative proxy for the business cycle. In both specifications, price dispersion, $DISP_{ijt}$, is measured in three different ways: the logarithm of the interquartile range, the Gini log-odds ratio,[6] and the 90th and 10th percentiles, each estimated in separate regressions. Analyzing the top and bottom of the price distribution separately provides additional information regarding the source of the change in price dispersion. Observations are weighted by the total number of passengers

[6]We measure price dispersion using the Gini log-odds ratio given by $G_{ij}^{lodd} = ln\left(\frac{G_{ij}}{1-G_{ij}}\right)$, which produces an unbounded statistic. No results change when the log of the Gini coefficient is used instead. See Hayes and Ross (1998) for further discussion.

on the route over the entire sample period and standard errors are clustered by route in order to control for autocorrelation as well as correlation between carriers on the same route. For robustness purposes, we ran specification (3.1) clustering by time period. This level of clustering accounts for any arbitrary correlation in the residuals by time period. Estimates of the coefficient on the output gap remain statistically significant at the 1 percent level.

3.3.1 Results

Table 3.2 contains estimation results for both specifications, using the logarithm of the interquartile range and the Gini log-odds ratio as the dependent variable. We report results for all direct routes in the 17-year sample.[7] The effect of a rise in the business cycle—as measured by the output gap—on price dispersion is positive and significant at the 1-percent significance level. The estimate indicates that a one percentage point rise in the output gap (i.e. from 0.01 to 0.02) is associated with an increase in the interquartile range by 1.56 percent and the Gini log-odds ratio by 0.011. The results from the second specification are similar to the first, indicating that a decrease in the unemployment rate is associated with an increase in the amount of price dispersion on a given route.[8] A one percentage point fall in the unemployment rate is associated with a 2.27 percent increase in the interquartile range.

A look at the estimates from the percentile regressions in Table 3.3 sheds further light on the manner in which price dispersion follows the business cycle. The estimates show that the output gap is positively correlated with the 90th-percentile price level but is not positively correlated with the 10th-percentile price level. An increase in the output gap by one percentage point is associated with a 1.16 percent increase in the 90th percentile price, but is not correlated with the 10th percentile price. Similarly, a fall in the unemployment rate by 1 percentage point is associated with a 1.37 percent increase in the 90th percentile

[7]This sample includes 154,407 carrier-route observations when using $\ln(IQR)$ as the dependent variable and 156,038 carrier-route observations using the Gini log-odds ratio. The reason for the slight decrease in the number observations is that observations in which the interquartile range was equal to zero were necessarily dropped.

[8]This sample includes 153,706 carrier-route observations when using $\ln(IQR)$ as the dependent variable and 155,331 carrier-route observations using the Gini log-odds ratio. We have fewer observations in this specification because we do not have unemployment information for American Samoa or St. Thomas.

price, while there is a statistically significant, but small -0.339 percent negative correlation between the unemployment rate and the 10th percentile price.

As in Gerardi and Shapiro (2009), we find that the effect of a decrease in competition—as measured by an increase in market concentration $\ln \widehat{HERF}$—on price dispersion is positive and significant at the 1-percent significance level. There also appears to be interesting dynamics occurring on the cost side. Fuel costs seem to filter into both the 10th percentile prices and 90th percentile prices, while the slower moving operating costs filter only into the 90th percentile prices. There are many plausible stories that could explain this result. One possibility may be that carriers simply find it easier to pass costs on to the more price-insensitive consumers as they are more likely to lose the more price-sensitive consumers to competition.

As an additional exercise, we split our sample between legacy carriers and low-cost carriers (LCCs). Legacy carriers tend to implement different pricing strategies compared to the LCCs, so it is important to assess whether the type of carrier plays an important role in how price dispersion varies with the business cycle. For instance, some legacy carriers offer "economy-plus," which offers passengers more leg room, separate access through security, and/or early boarding. To determine whether these different types of carriers actually price differently over the business cycle, we re-estimate the main econometric specification for each sample separately. The estimates divided by carrier type are reported in the Table 3.4 and show that most of the effects from the business cycle on price dispersion in the full sample of routes stem from the legacy carriers. The effect of the output gap on the interquartile range is slightly larger than two times the magnitude in the sample of legacy carriers ($\hat{\beta}_1$ = 2.335 compared to the estimated effect in the sample of LCCs $\hat{\beta}_1$ = 0.965).

Overall, the fixed-effects, panel estimates provide evidence of a positive relationship between the business cycle and price dispersion in the airline industry. Furthermore, the results show that the pro-cyclicality of price dispersion is largely driven by prices near the top of the price distribution.

3.4 Discussion

In this section, we discuss some potential explanations for our empirical findings. Because the BTS does not provide detailed information on specific ticket or demographic characteristics, we cannot unequivocally single out any one specific pricing mechanism. However, we are able to provide empirical and theoretical evidence that favors certain explanations over others. We focus on two widely discussed theories of price dispersion in the airline industry: price discrimination and stochastic-demand pricing.

3.4.1 Second-Degree Price Discrimination

The practice of price discrimination is one of the leading explanations for price dispersion in the airline industry. Airlines implement price discrimination techniques by segmenting heterogeneous groups of consumers and charging them distinct prices for a homogeneous product. Advance purchase requirements, non-refundable tickets, and Saturday-night layovers are a few examples of restrictions that airlines use to identify passengers with different price elasticities of demand. Since high-income or business consumers tend to place a high value on their time, they are more likely to purchase more expensive tickets without such restrictions. By making use of these techniques, airlines are able to separate price-sensitive travelers from price-insensitive travelers.

Using a parsimonious framework of second-degree price discrimination, we illustrate below that pro-cyclical price dispersion may be a side-effect of second-degree price discrimination. Specifically, under plausible assumptions of the utility function, a price discriminatory pricing policy implies that prices in the upper tail of the price distribution will be more sensitive to aggregate income fluctuations than prices in the lower tail. This suggests that price dispersion will positively covary with aggregate income when firms are price discriminating between consumers with different willingness-to-pay.

Consumers

Consider a simple model where consumers differ in their level of income, y. A consumer solves the following constrained utility maximization problem:

$$\max_{d\in\{0,1\}} d \cdot x + u(m) \tag{3.3}$$

subject to:

$$y = m + d \cdot p$$

where x represents the valuation of the ticket, which for now we assume is a constant. The variable m represents the numeraire commodity and d represents the consumer's decision to buy or not buy the good. Note that $u(\cdot)$ is the functional form representing the manner in which the consumer values the numeraire commodity relative to the discrete good, and we assume that it displays the conventional properties: $u'(y) > 0$ and $u''(y) < 0$. It follows that the indirect utility function for the case in which the consumer purchases the discrete good ($d = 1$) is given by:

$$U = x + u(y - p). \tag{3.4}$$

As in Tirole (1988), we make the assumption that a consumer's income is very large relative to the consumer's valuation, x, and subsequently to the equilibrium price charged. This allows us to take a first-order Taylor expansion around $p^* = 0$, which under the assumption that $y - p \approx y$, yields:

$$U = x + u(y) - u'(y)p. \tag{3.5}$$

It follows that for a given consumer to be better off consuming the good, it must be the case that $x + u(y) - u'(y)p \geq u(y)$, which means demand for the good is:

$$d(p) = \begin{cases} 1 & \text{if} \quad p \leq \frac{x}{u'(y)} \\ 0 & \text{if} \quad p > \frac{x}{u'(y)} \end{cases} \tag{3.6}$$

Firm behavior

To simplify the firm's problem we assume two types of consumers and two types of tickets. The results below can be easily generalized to N types of consumers and N types of tickets. We assume there exists a share α of high income consumers with income y_h and a share $1-\alpha$ of low income consumers with income y_l. Quality is indexed by v, and we assume that there exists a high quality ticket, x_1, and a low quality ticket, x_2. For instance, $v = 2$ indicates a ticket that has an advance purchase requirement or Saturday-night stayover requirement, while $v = 1$ indicates a less restrictive ticket. It follows that with positive time costs, the net quality of $v = 1$ will be higher than that of $v = 2$ such that $x_1 > x_2$.

The firm's problem in the two-consumer-type case is to maximize profits given consumer demand derived above. The firm has the option to separate the market by offering different types of tickets. To obtain a separating equilibrium, the firm must be able to separate the market and also find it profit-maximizing to do so. It follows from Mussa and Rosen (1978) and Tirole (1988) that optimal incentive-compatible prices satisfy:

$$p_1^* = b_h x_1 - (b_h - b_l)x_2 \tag{3.7}$$

$$p_2^* = b_l x_2. \tag{3.8}$$

where $b_h = \frac{1}{u'(y)|_{y=y_h}}$ and $b_l = \frac{1}{u'(y)|_{y=y_l}}$. As the high-income consumer values x_2 more than the low-income consumer, the firm must lower the price of x_1 to dissuade the high-income consumer from deviating and purchasing the lower quality ticket, x_2. Specifically, the price is lowered by the extra utility the high-income consumer would have received over the low-income consumer by consuming x_2, $(b_h - b_l)x_2$. This lower price ensures that the high-income consumer does not purchase x_2 instead of x_1 (this ensures that the equilibrium is incentive compatible).[9]

[9]Maskin and Riley (1984) deal with a more general case where there is a choice over both quality and quantity. The authors show that the optimal quantity level is a function of the underlying quality. This implies that when consumers are allowed to make choices over both quantity and quality, it is optimal for the monopolist to offer a price schedule such that for a given quality level, he offers a unique quantity level associated with it, so that in fact consumption choices are made discrete. Hence *vis à vis* our discrete choice framework, the necessary correction would be taking into account the possibility of discrete quantity differences between the two consumer types.

Price sensitivities to a change in income, y, will be:

$$\frac{\partial p_1^*}{\partial y} = A_h b_h(x_1 - x_2) + A_l b_l x_2 \tag{3.9}$$

$$\frac{\partial p_2^*}{\partial y} = A_l b_l x_2 \tag{3.10}$$

where $A_h = \left[-\frac{u''(y)}{u'(y)}\right]_{y=y_h}$ and $A_l = \left[-\frac{u''(y)}{u'(y)}\right]_{y=y_l}$ are Arrow-Pratt measures of absolute risk-aversion (ARA) evaluated at income levels y_h and y_l, respectively. As the ARA will be positive as long as consumers have diminishing marginal utility of income, it follows from (3.1) and (3.2) that prices at the upper end of the distribution will be more sensitive to income shocks than prices in the bottom portion of the distribution. By contrast, if the firm chooses a uniform pricing strategy and wishes to sell to all consumers (i.e. both high- and low-income consumers), it must set a price in accordance with the low-income consumer's preference parameter, b_l. For instance, if the firm chooses to sell only the low-quality ticket, it would set a uniform price $b_l x_2$. If it chooses to sell only the high-quality ticket it would set a uniform price $b_l x_1$. In the former case, average price sensitivity under a uniform pricing strategy will be less than the average price sensitivity under a discriminatory pricing strategy as long as $\alpha > 0$. In the latter case, this will be true only if $\alpha > \frac{A_l b_l}{A_h b_h}$.

The price range between the high and low price ticket (a measure of price dispersion similar to the interquartile range) is:

$$D = p_1 - p_2 = b_h(x_1 - x_2). \tag{3.11}$$

It follows that the elasticity of price dispersion relative to a change in aggregate income, y, is:

$$\varepsilon_{D,y} = \left[\frac{\partial D}{\partial y}\frac{y}{D}\right]_{y=y_h} = \left[-\frac{u''(y)}{u'(y)}y\right]_{y=y_h} \tag{3.12}$$

which is simply the coefficient of relative risk aversion (CRA) evaluated at y_h. As long as the CRA is positive (i.e. diminishing marginal utility of income), price dispersion will widen with an increase in aggregate income. It is important to note that this result is also sensitive to the choice of the utility function. Specifically, it will only hold for utility functions with the property that the consumer's willingness-to-pay increases jointly with both quality and income.

The model also shows that second-degree price discrimination may cause pro-cyclical price dispersion for reasons other than relative movements in price elasticities. For instance, equation (3.11) implies that price dispersion will follow the business cycle if there are relative movements in the non-price attributes of the good, $x_1 - x_2$, over the business cycle. This could happen if time-costs are cyclical, or more generally speaking, if there are complementarities with certain ticket characteristics and business cycle conditions. Overall, this framework shows that price discriminatory tactics can cause price dispersion to widen during economic booms due to (1) movements in price elasticities or (2) movements in the non-price attributes of the good (that is, x). Since the DOT data does not include many of the ticket characteristics and because we do not have demographic information of the ticket purchasers, we cannot distinguish between these two effects.

An Empirical Exercise

To address the role of price-discriminatory behavior in generating pro-cyclical price dispersion, we perform an additional empirical exercise. Specifically, we assess the impact of consumer heterogeneity on the cyclicality of price dispersion by estimating equations (3.1) and (3.2) on two subsamples of routes: a sample of routes that are characterized by significant consumer heterogeneity in willingness-to-pay, as well as a sample of routes in which there is a more homogeneous consumer base. *Ceteris paribus*, there should be more opportunities to price discriminate in the former sample due to the presumed larger difference in willingness-to-pay of the consumers. To be clear, this is not a formal test of the theoretical model, but rather an additional specification of our empirical analysis of Section 3.1.

We decompose the full sample of routes into "big-city" routes and "leisure" routes, which we believe correspond to markets with heterogeneous consumer bases and markets with more homogeneous consumer bases, respectively. Since routes between large cities tend to attract both business and leisure travelers, they tend to have a bimodal distribution of prices while routes to largely leisure destinations, such as islands and beaches, tend to have unimodal price distributions and lower median prices.[10] Thus, airlines may have more opportunities to implement price discrimination strategies on these big-city routes since they include relatively more high-income, business consumers. Furthermore, note that equation (3.12) implies that dispersion on big-city routes will be more sensitive to the cycle if the utility over income displays increasing relative risk aversion and less sensitive to the cycle if it displays decreasing relative risk aversion.[11] Table 3.1 shows summary statistics for the explanatory variables in each of these two samples. The table shows that costs are similar between the two samples, however, price dispersion is much larger in the sample of big-city routes.

Tables 3.5 contains estimates of the correlation between the interquartile range of the price distribution for a carrier-route observation and the business cycle for big-city routes versus leisure routes. The estimates show that price dispersion is more closely tied with the output gap and the average city-wide unemployment rate for the big-city route sample than the leisure sample. For instance, in the big-city route sample, a one percentage point rise in the output gap is associated with a 2.7 percent increase in the interquartile range, while it is associated with a statistically insignificant 0.19 percent increase in the leisure route sample. The effect of the unemployment rate on price dispersion is slightly larger in the big-city sample, a coefficient of 2.8 compared to 2.3 in the leisure sample. This smaller difference in magnitude of the coefficients may be due to the fact that, as opposed to the output gap, the unemployment rate is specific to the economic conditions at the endpoint cities.

3.4.2 Stochastic-Demand Pricing

Another important theory regarding the existence of price dispersion is that of stochastic-demand pricing. If the carrier is constrained by capacity, then as more flights reach full capacity, the expense of an additional passenger becomes very large as either a bigger aircraft or an extra flight is needed to supply the extra seat-mile. Eden (1990) shows that effect can induce price dispersion to rise in periods of peak demand when full capacity is

[10] For a full list of the cities in each sample as well as a detailed description of how these subsamples are created see Gerardi and Shapiro (2009).

[11] This can be seen by taking the derivative of (3.12) with respect to y_h.

reached.

In discussing the effect of capacity constraints on pricing, it is useful to decompose marginal cost into its two primary components, which we refer to as the passenger cost and the capacity cost. If the aircraft is not operating at full capacity, then marginal cost is simply equal to the passenger cost; the cost of adding an additional passenger to the airplane. This cost is mostly made up of the extra fuel required to transport the additional weight of the passenger, while other, lesser components include the in-flight costs of serving the additional passenger (i.e. meals, snacks, etc.). However, if the airplane is operating at full capacity, then marginal cost is equal to the direct cost of an additional passenger as well as the more substantial cost of an additional flight. This cost is incurred regardless of whether or not seats on the airplane are filled with passengers, while the passenger cost is only incurred on seats that are sold. This implies that marginal cost at the route level is given by,

$$c_{ij} = \begin{cases} \beta_{ij}, & \text{if capacity is not reached} \\ \beta_{ij} + \lambda_{ij}, & \text{if capacity is reached} \end{cases}$$

where β_{ij} is the cost of serving an additional passenger one mile on route j by carrier i, and λ_{ij} is the cost of an additional flight (in seat-miles).

If airlines account for stochastic demand concerns in their pricing decisions, then aggregate demand fluctuations could alter a firm's expected probability of selling a ticket, and subsequently alter the "effective" capacity cost. In particular, if ex-ante the carrier is uncertain about the level of demand for a flight, then under price-setting commitments and costly capacity, profit-maximizing behavior induces a distribution of prices rather than a single price. The intuition is that if the firm were allowed to change price after the realization of the state, then it would set a low price in the low-demand state and a high price in the high-demand state. However, because the firm must commit to a menu of prices ex-ante, its profit maximizing strategy is to assign multiple prices to specified quantities of the good. That is, if a firm must pay costs irrespective of whether or not its output is sold, then it has a large incentive to set higher prices on goods that are less likely to be sold.

Eden (1990) formalized a model in a setting of perfect competition where there is uncertainty regarding the number of agents who will show up to exchange goods in the marketplace. In such a setting, goods are characterized by the probability that they will be sold, and in equilibrium firms face a trade-off between price and the probability of sale. In the model, equilibrium prices are given by the condition,

$$p_s = \beta + \underbrace{\frac{\lambda}{prob(sale)_s}}_{\lambda_s^{eff}} \tag{3.13}$$

where p_s is the price of the sth good, β is an operating cost that the firm must pay for each good that it sells, λ is the unit capacity cost, and $prob(sale)$ is the probability that good s is sold. The second term on the right-hand side of the equation can be interpreted as an "effective" capacity cost of good s, λ_s^{eff}. This term implies that in competitive equilibrium, firms are indifferent between selling a high-priced good with low probability and selling a low-priced good with high probability. Dana (1999) extended Eden's model to monopoly and oligopoly market structures. With stochastic demand, the monopolist sets a higher price for a good that sells only in high demand states since its effective cost is higher.

In this setting, when the carrier commits to prices ex-ante, the highest priced tickets—tickets with the highest effective capacity cost—are not purchased until demand rises sufficiently high to purchase all of the low priced tickets. Thus, if the carrier is pricing solely with stochastic demand concerns, then peaks in aggregate demand will induce higher price dispersion through the higher effective capacity cost of the remaining seats on crowded aircraft.

An Empirical Exercise

To asses the empirical importance of stochastic-demand pricing in generating pro-cyclical price dispersion we exploit the expected relationship between capacity utilization and price dispersion that would arise if stochastic demand played an important role in airline pricing tactics. Specifically, under stochastic-demand pricing, utilization should positively co-vary with price dispersion because high-priced tickets would be purchased only when aircraft are near full capacity.

Figure 3.4 shows the mean aircraft capacity utilization rate over the sample period. Interestingly, utilization has been steadily increasing over the course of the sample period, fluctuating with some seasonal variation. As a formal test, we control for the effects of stochastic-demand pricing on price dispersion by including a measure of carrier i's utilization rate on route j in our estimation routine, $UTIL_{ijt}$. As this variable is potentially endogenous, we also include specifications where we use instrumental variables. Specifically, we instrument for $UTIL_{ijt}$ using the utilization of route j in period t, $UTIL_{jt}$. This variable should be correlated with airline i's specific utilization rate due to variations in aggregate demand for route j. As Figure 3.4 makes apparent, it may be important to remove low frequency components from the utilization variable. Thus, we also include specifications with a de-trended measure of aircraft utilization, $UTIL_{dt}$.[12]

We report results of this exercise using two measures of price dispersion. Estimates using the logarithm of the interquartile range are reported in Table 3.6 and estimates with the Gini-log odds ratio are reported in Table 3.7. In all specifications, the coefficient on our measure of the business cycle is positive and statistically significant at the one-percent level. Thus, holding fixed aircraft utilization, price dispersion remains pro-cyclical. These estimates suggest that the pro-cyclicality of variation in airline price dispersion is likely not tied to variation in capacity utilization. In turn, this suggests that stochastic-demand pricing strategies do not explain our findings. While this analysis favors price discrimination as the explanation for the pro-cyclical nature of airline price dispersion, it is important to stress that we can only favor certain explanations over others due to certain limitations of the data. For instance, we use a monthly measure of capacity utilization at the carrier-route level, whereas ideally, we would like capacity utilization measured at the flight level. It is comforting to note, however, that our results correspond with recent work of Puller et al. (2009) who use more granular ticket information.

[12]The measure $UTIL_{dt}$ is measured by collecting the residuals from running a regression of utilization, $util$, on a cubic-polynomial time trend and quarter dummies.

3.5 Conclusion

In this paper we have documented that price dispersion is significantly pro-cyclical in the airline industry. We show that the empirical results are consistent with a parsimonious discrete-choice model of second-degree price discrimination. In addition, we implement a few empirical exercises that provide support for this interpretation over others such as stochastic demand pricing or pro-cyclical variation in airline costs. With the available data, we cannot completely rule out other mechanisms that could create pro-cyclical behavior in price dispersion. One such mechanism is changes in consumer behavior over the business cycle. Even in the absence of variation in airlines' pricing strategies over the business cycle, if consumers simply purchase more expensive airline tickets with fewer restrictions in boom periods and less expensive tickets with more restrictions in more depressed periods, then our measured degree of price dispersion would be expected to co-vary with the business cycle. Future research on this topic will hopefully employ more detailed data that will have the ability to distinguish between causal mechanisms, and thus determine if price discrimination is the most important driver of price dispersion at the business cycle frequency. Another potentially fruitful avenue for future research is to analyze price dispersion in other industries. If pro-cyclical price dispersion is largely due to price discrimination tactics, then we would expect to find pro-cyclical price dispersion in industries in which firms rely on price discriminatory strategies, such as hotels, stadiums, restaurants, theaters, yellow-page advertising, cement, and personal computers.

Another interesting extension of this study would be to assess whether price discriminatory tactics act to accentuate the degree to which airline profits fluctuate over the business cycle. Given the high volatility of profits over the course of the last two decades as well as the large number of bankruptcies by legacy carriers, the airline industry seems particularly sensitive to aggregate demand conditions. But while legacy carriers have struggled, the LCCs have somehow managed to stay profitable during this era. One possibility is that the large profit swings of legacy carriers relative to LCCs are, in part, attributable to differences in the reliance on price discriminatory tactics, as LCCs such as Southwest and JetBlue do not price discriminate to the same extent as legacy carriers.

3.6 Appendix

Variable Definitions

- ln$P(k)_{ijt}$ - The logarithm of the kth price percentile of carrier i on route j in period t, obtained from the DB1B.

- $\ln IQR_{ijt}$ - The logarithm of the interquartile range, given by $P(75)_{ijt}$ - $P(25)_{ijt}$, where $P(k)_{ijt}$ is the price percentile of carrier i on route j in period t, obtained from the DB1B.

- $Gini_{ijt}^{lodd}$ - The Gini log-odds ratio, given by $G_{ijt}^{lodd} = \ln(\frac{G_{ij}}{1-G_{ij}})$, where G_{ijt} is the Gini coefficient of carrier i's price distribution on route j in period t, calculated using data from DB1B.

- ln$HERF_{jt}$ - The logarithm of the Herfindahl-Hirschman index of route j in period t, calculated using passenger shares obtained from the DB1B.

- $YGAP_t$ - The log of nominal GDP in period t minus the log of nominal potential GDP in period t, as measured by the Congressional Budget Office (CBO).

- UR_{jt} - The average metropolitan unemployment rate in period t of the origin and destination state of route j, obtained from Bureau of Labor Statistics (BLS).

- $\ln FUEL_{it}$ - The average cost per gallon fuel by carrier i in period t, obtained from the BTS P-52 database.

- $\ln COST_{it}$ - Total operating costs minus total fuel costs divided by total seat-miles for carrier i in period t, obtained from the BTS P-52 database.

- $UTIL_{ijt}$ - The capacity utilization rate of carrier i on route j in period t measured by total passengers divided by total seats. Obtained from the T-100 database.

- $UTIL_{dt}$ - The de-trended capacity utilization rate of carrier i on route j in period t measured as the residual from the regression of $util_{ijt}$ on a cubic-polynomial time trend and quarter dummies.

Instruments

- ln$PASSRTE_{jt}$ - The logarithm of total enplaned passengers on route j in period t from the T-100 Domestic Segment Databank.

- $IRUTHERF$ - This instrument is identical to one used by Borenstein and Rose (1994). This variable is the square of the fitted value for $MKTSHARE_{ijt}$ from its first-stage regression, plus the rescaled sum of the squares of all other carrier's shares. See Borenstein and Rose (1994) for a more detailed explanation. It is equal to $\widehat{MKTSHARE_{ijt}}^2 + \frac{HERF_{jt} - MKTSHARE_{ijt}^2}{(1-MKTSHARE_{ijt})^2} * (1 - \widehat{MKTSHARE}_{ijt})^2$.

- $GENSP$ - $\frac{\sqrt{ENP_{j1} * ENP_{j2}}}{\sum_k \sqrt{ENP_{k1} * ENP_{k2}}}$, where k indexes all airlines, j is the observed airline, and ENP_{k1} and ENP_{k2} are airline k's average quarterly enplanements at the two endpoint airports. This instrument is similar to one used by Borenstein and Rose (1994), with the difference being that Borenstein and Rose use average daily enplanements, while we use average quarterly enplanements, as a result of data availability. Data on enplanements were obtained from the T-100 Domestic Segment Databank.

Data Construction

In this appendix, we discuss our methods and assumptions involved in constructing our panel of airline-route ticket observations from the DB1B and T-100 Domestic Segment databases maintained by the BTS at their online website, Transtats. There are three different sub-components to the DB1B data set. They are market data, coupon data, and ticket data, and we combine variables from all three sources.[13]

We use only domestic, coach-class itineraries and keep only tickets containing direct flights.[14] Direct flights typically account for 30 percent of the itineraries in the DB1B over the course of our sample, with no apparent trend.

The BTS includes a variable that describes the reliability of each ticket price ("dollar cred"). The variable takes on a value of 0 if the fare is of questionable magnitude, based

[13] For further reference, see the BTS's website http://www.transtats.bts.gov.

[14] The sample of direct flights encompasses both non-stop flights and flights in which there is a stop but no change of plane.

on a set of limits defined by the BTS, and it takes a value of 1 if it is credible. We drop all tickets for which this variable takes a value of zero.

The DB1B also provides limited information regarding the fare class of each ticket. Each ticket is labeled as either coach-class, business-class, or first-class, and we eliminated all first-class and business-class itineraries. Unfortunately, the DB1B does not have any direct way of identifying frequent-flyer tickets, but there are indirect methods that have been used in the previous literature, and we follow these in our analysis. First, we drop all fares coded as 0. Next, we dropped all fares that are less than or equal to $20 ($10 for one-way tickets).

In addition to eliminating frequent-flyer tickets and higher-class tickets, we also eliminate tickets in which the operating and ticketing carriers are different due to code sharing arrangements. Code sharing is a practice where a flight operated by an airline is jointly marketed as a flight for one or more other airlines. Due to the uncertainty regarding the actual airline who is setting the price schedule in such an arrangement, we decided to eliminate these itineraries. Code sharing first appears in the data in 1998:Q1. On average, approximately 80 percent of the original number of direct tickets in the DB1B is retained in the analysis.

After filtering the ticket data for each quarter of the DB1B, we combined tickets from all 55 quarters and collapsed the data into airline-route observations. For example, if we had 10,000 United Airline tickets between Boston and Los Angeles in 1993:Q1, we calculated summary statistics (such as the Gini coefficient), and collapsed the data into a single observation corresponding to a United Airlines flight between Boston and Los Angeles in 1993:Q1.

The merge between the DB1B and T-100 Segment databases was not exact (around 45 percent matched). First, since the DB1B does not provide complete coverage for all airlines and routes, there are a number of direct routes in the T-100 data that we do not find in the DB1B (especially low-volume routes). Second, the DB1B does not allow us to distinguish between a non-stop, direct ticket and a ticket that involves a stop without a plane change. For example, if a passenger takes a flight from Boston to Orlando that stops in Atlanta, but does not involve a plane change, his itinerary will look identical to that of a passenger who flies from Boston to Orlando without any stops. For this reason, we identified some airline routes as direct in the DB1B, that are not non-stop, and therefore do not have segment information in the T-100 data. While we lose many airline-route observations during the merge as a result, we believe that this merge actually provides a nice filter, since we would ideally like to use only non-stop, direct flights. Thus, by merging data between the DB1B and the T-100, we likely eliminate a large proportion of flights that are direct, but not non-stop due to a plane change.

In an effort to eliminate possible coding errors, we drop certain airline-route observations from the data that we believe do not have adequate coverage to calculate reliable price dispersion statistics. We drop any airline-route observation that does not have at least 100 passengers in the DB1B. Furthermore, for each airline route observation, we calculate the average number of passengers over time in both the DB1B and the T-100 Segment databases. If the number of passengers on an airline route in a given quarter falls below 25 percent of its mean over time in one of the databases, but not in the other, then we drop the observation from our data, on the basis that its value is most likely measurement error. However, if the number of passengers on an airline route in a given quarter falls below 25 percent of its mean in both the DB1B and the T-100 Segment databases, then we keep the observation in our data.

Finally, we addressed the issue of "double counting." Since we defined a route as a directional trip in our data, any round-trip ticket would count twice. For example, a round-trip fare from Boston to San Francisco would appear twice in the data — once as BOS-SFO and once as SFO-BOS. Since this would have no effect on the consistency of our estimates, but a significant effect on the size of our standard errors, we chose to drop one of the directions. Of course, the drawback of this assumption is that some one-way fares were dropped from the data as a result. In our judgment, the first issue outweighed the second issue.

Figures

Figure 3.1: Price Dispersion over the Business Cycle

Depicted as a grey line is the passenger-weighted average of the interquartile range for all routes in the DB1B database. The solid black line is the five-quarter moving average. The output gap, as measured by the Congressional Budget Office (CBO), is also depicted as a dashed red line.

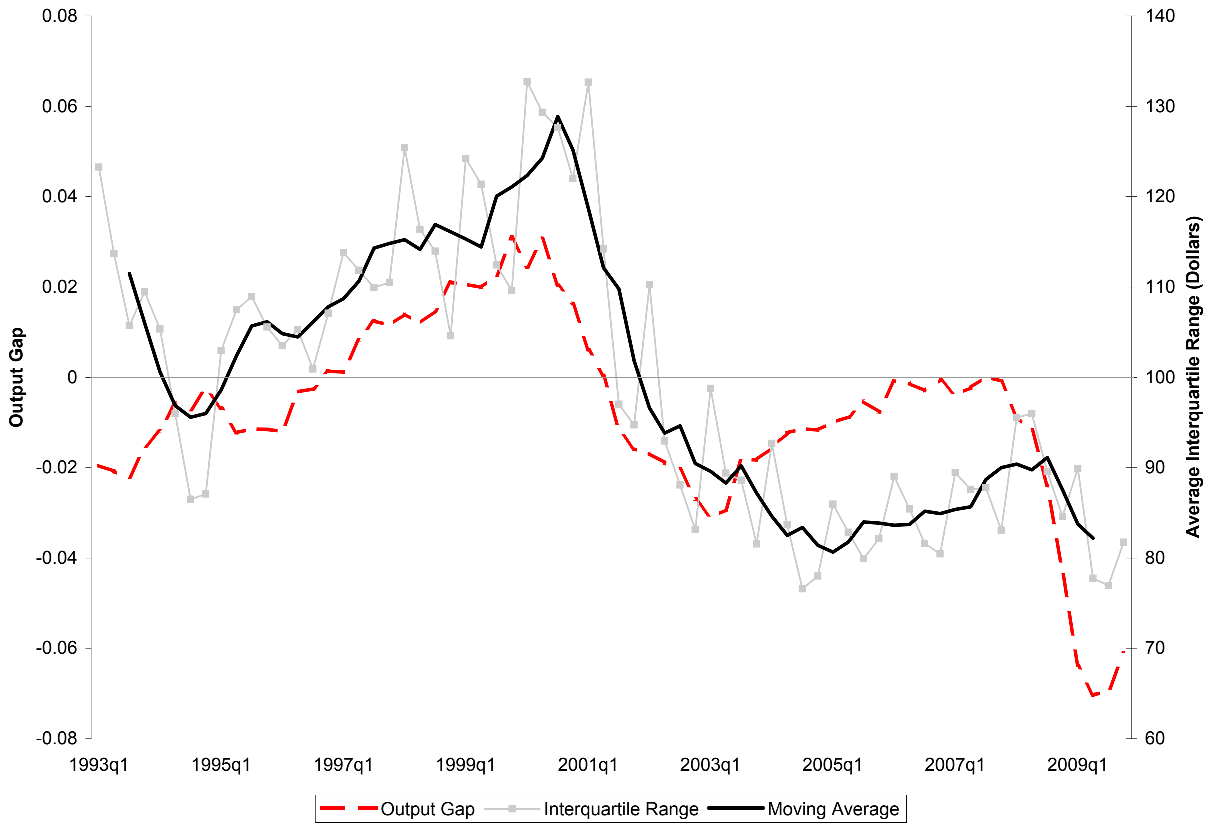


Figure 3.2: Pricing Dynamics

Depicted are 10th, 30th, 50th, 70th and 90th price percentiles for three airline-route observations. The output gap, as measured by the Congressional Budget Office (CBO), is also depicted as a dashed line.

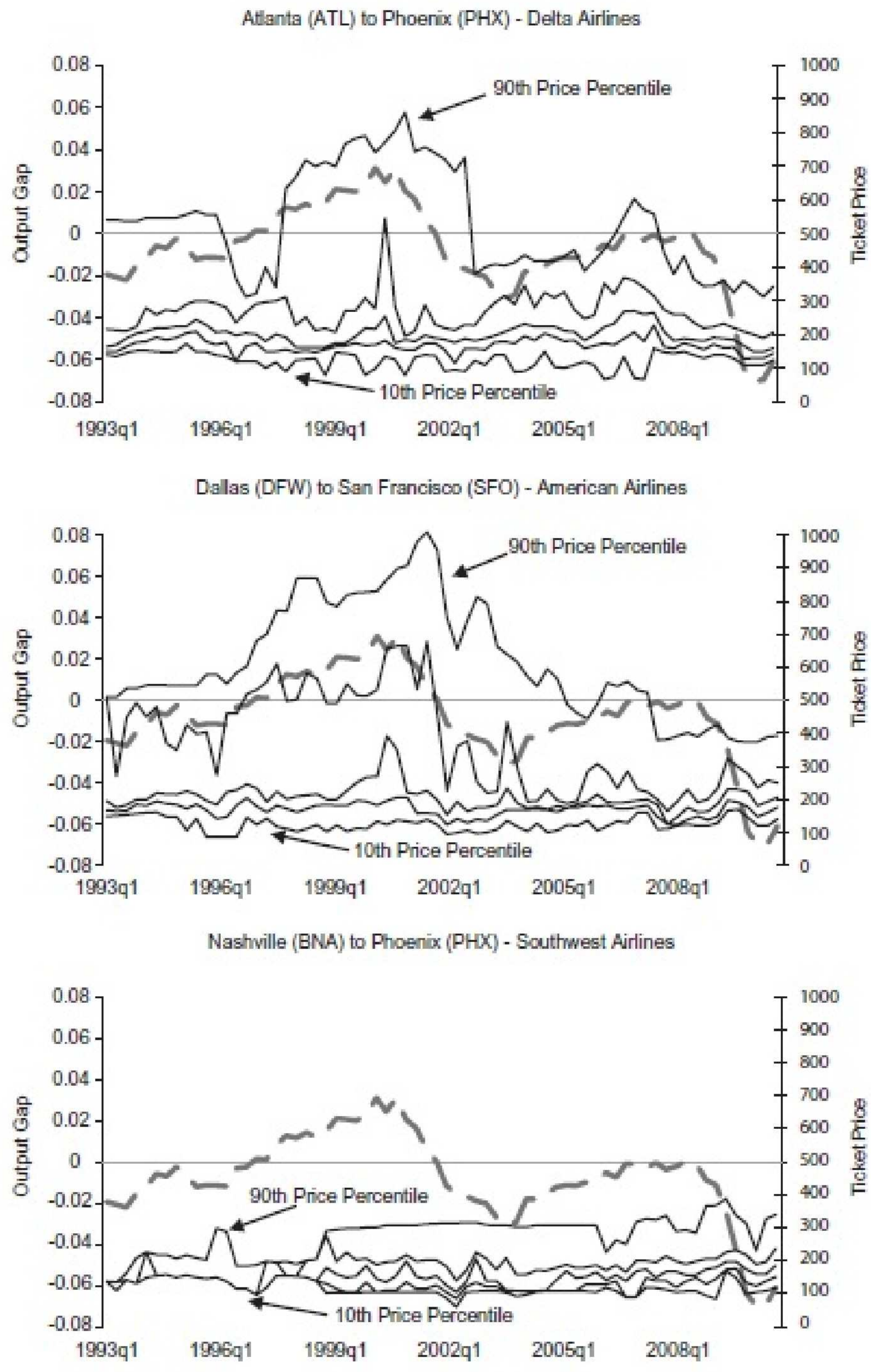


Figure 3.3: Aircraft Operating Costs

Depicted are total aircraft operating cost (including fuel) in dollars per seat-mile for four carriers: US Airways, United Airlines, JetBlue, and Southwest.

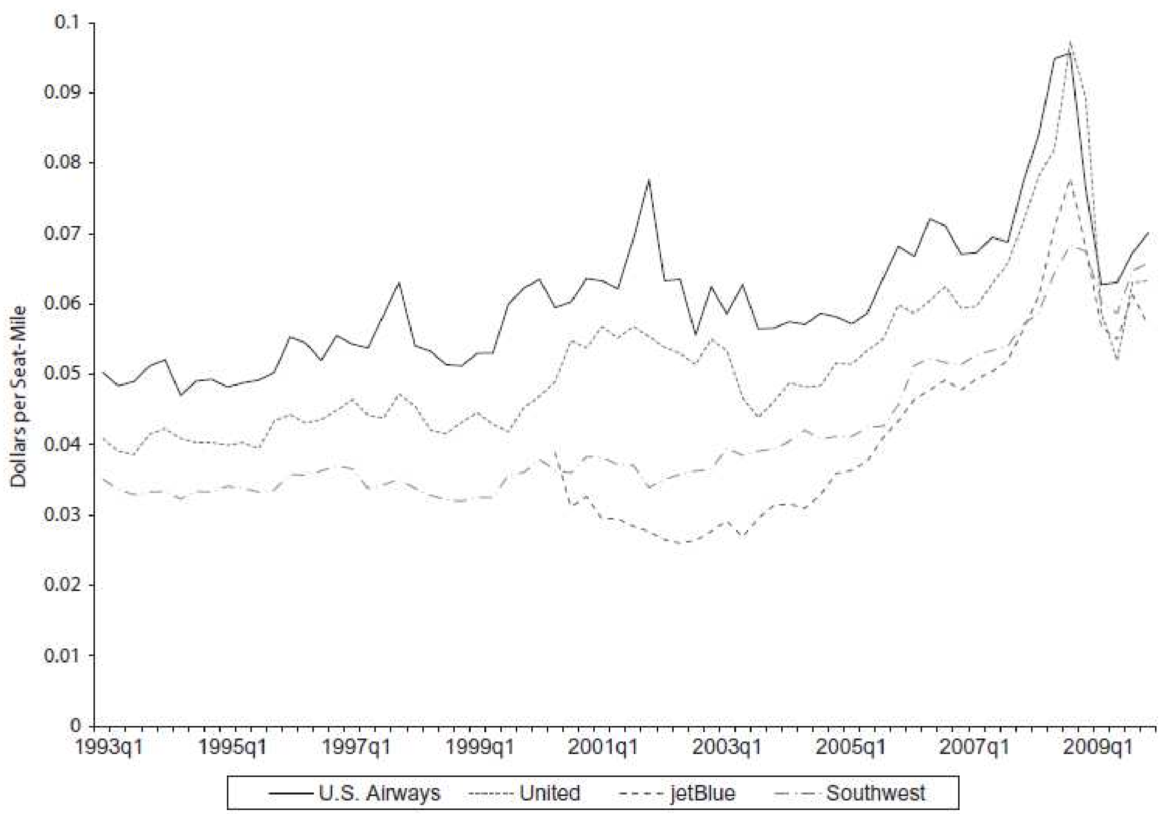
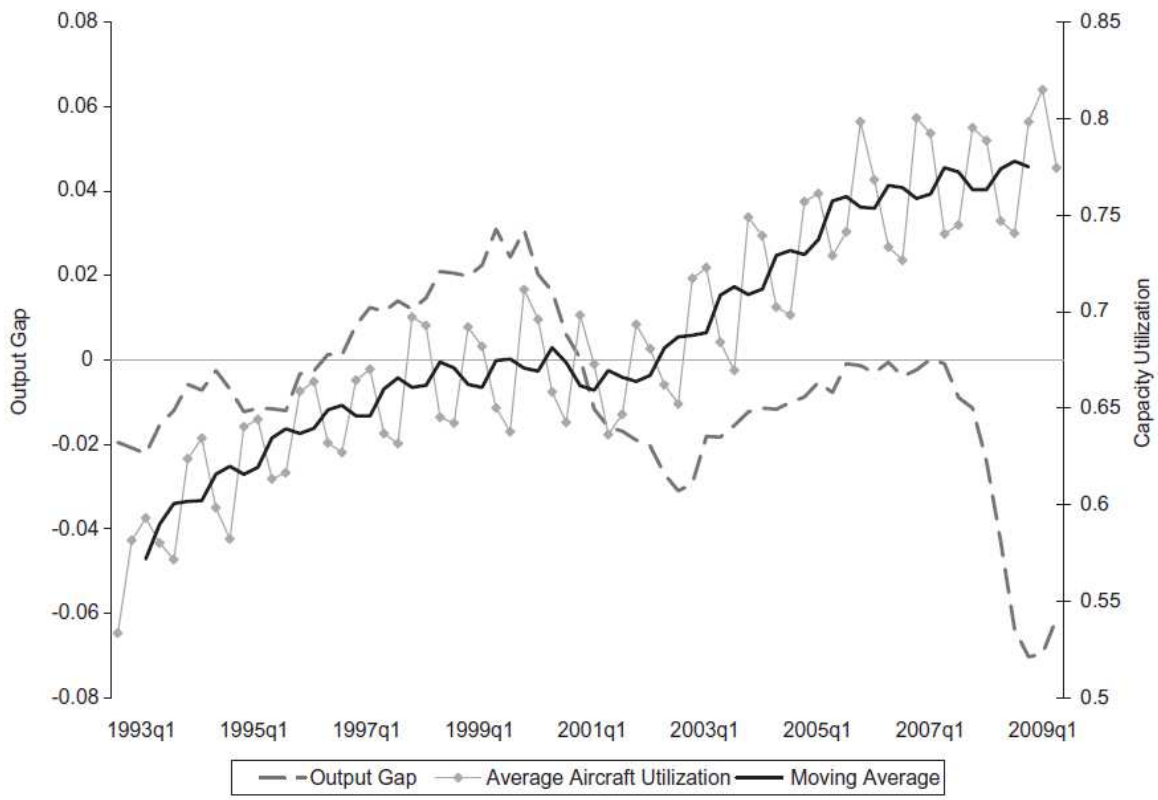


Figure 3.4: Aircraft Utilization Rate

Depicted as a grey line is aircraft utilization, measured as total passengers divided by total available seats from the BTS T100 database, for each quarter in the sample period. The black line depicts a moving average of this measure. The output gap, as measured by the Congressional Budget Office (CBO) is depicted as a dashed line.



Tables

Table 3.1: Summary Statistics

The interquartile range (IQR), 90th percentile price (90th Pctl.), 10th percentile price (10th Pctl.), and fuel cost per gallon (FUEL) are reported in dollars. Total operating fuel cost less fuel per seat-mile (COST) is reported in cents.

	Full Sample		Legacy		LCC		Big-City		Leisure	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Gini	0.22	0.08	0.25	0.07	0.17	0.05	0.25	0.08	0.19	0.07
IQR	92	83	112	92	52	27	117	105	64	51
90th Perc. Price	278	161	327	170	178	70	338	201	232	125
10th Perc. Price	95	39	102	40	80	29	99	37	98	48
HERF	0.76	0.25	0.75	0.25	0.77	0.25	0.67	0.24	0.77	0.25
COST	3.25	0.81	3.37	0.59	2.83	0.52	3.27	0.65	3.21	0.88
FUEL	1.26	0.98	1.15	0.82	1.45	0.76	1.24	0.84	1.41	1.36
UTIL	0.69	0.15	0.69	0.15	0.69	0.15	0.69	0.15	0.73	0.15

Table 3.2: Full Sample Estimates

All regressions include carrier-route-specific dummies and quarter dummies. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

	$ln(IQR)$		$Gini^{lodd}$	
YGAP	1.561***		1.145***	
	(0.199)		(0.104)	
-UR		2.271***		1.641***
		(0.286)		(0.152)
ln $\widehat{HERF}$	0.261***	0.258***	0.079***	0.076***
	(0.043)	(0.043)	(0.023)	(0.023)
ln FUEL	0.046***	0.034***	-0.042***	-0.052***
	(0.015)	(0.016)	(0.009)	(0.009)
ln COST	0.135**	0.114**	0.450***	0.436***
	(0.061)	(0.057)	(0.030)	(0.030)
Observations	154407	153706	156038	155331

Table 3.3: Full Sample Estimates: Percentiles

All regressions include carrier-route-specific dummies and quarter dummies. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

	$ln(90)$		$ln(10)$	
YGAP	1.157***		-0.031	
	(0.104)		(0.061)	
-UR		1.373***		0.339***
		(0.151)		(0.105)
ln $\widehat{HERF}$	0.299***	0.296***	0.218***	0.218***
	(0.029)	(0.029)	(0.015)	(0.014)
ln FUEL	0.034***	0.024***	0.097***	0.097***
	(0.011)	(0.012)	(0.005)	(0.005)
ln COST	0.292**	0.276**	-0.015	-0.016
	(0.034)	(0.033)	(0.016)	(0.015)
Observations	156038	155331	156038	155331

Table 3.4: Panel Estimates by Carrier Type

The dependent variable is the logarithm of the interquartile range. All regressions include carrier-route-specific dummies and quarter dummies. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10 per cent, 5 per cent, or 1 per cent significance level, respectively.

	Legacy		LCC	
YGAP	2.335***		0.965***	
	(0.224)		(0.175)	
-UR		2.677***		1.747***
		(0.317)		(0.276)
ln $\widehat{HERF}$	0.191***	0.185***	0.311***	0.315***
	(0.048)	(0.048)	(0.039)	(0.039)
ln FUEL	-0.006	-0.027	0.225***	0.215***
	(0.016)	(0.016)	(0.013)	(0.012)
ln COST	0.069	0.029	0.579***	0.597***
	(0.045)	(0.045)	(0.042)	(0.043)
Observations	105636	104994	40941	40926

Table 3.5: Panel Estimates by Route Type

The dependent variable is the logarithm of the interquartile range. All regressions include carrier-route-specific dummies and quarter dummies. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10 per cent, 5 per cent, or 1 per cent significance level, respectively.

	Big-City		Leisure	
YGAP	2.732***		0.194	
	(0.352)		(0.467)	
-UR		2.756***		2.248***
		(0.482)		(0.605)
ln $\widehat{HERF}$	0.334***	0.329***	0.195***	0.209***
	(0.078)	(0.079)	(0.048)	(0.053)
ln FUEL	-0.054**	-0.080**	0.190***	0.194***
	(0.024)	(0.026)	(0.027)	(0.031)
ln COST	0.093	0.047	-0.085	-0.075
	(0.070)	(0.070)	(0.162)	(0.151)
Observations	43614	43614	35312	34611

Table 3.6: Estimates with Capacity Utilization Control: IQR

All regressions include carrier-route-specific dummies and quarter dummies. IV estimates indicate that UTIL and UTILdt were instrumented using utilization at the route-quarter level. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10 per cent, 5 per cent, or 1 per cent significance level, respectively.

$ln(IQR)$								
	OLS				IV			
YGAP	1.896***	2.161***			1.923***	2.196***		
	(0.213)	(0.217)			(0.216)	(0.227)		
-UR			2.687***	3.007***			2.721***	3.055***
			(0.281)	(0.287)			(0.279)	(0.275)
$UTIL$	-1.122***		-1.121***		-1.211**		-1.213***	
	(0.084)		(0.081)		(0.101)		(0.096)	
$UTIL_{dt}$		-1.217***		-1.208***		-1.286***		-1.287***
		(0.077)		(0.0703)		(0.109)		(0.104)
$ln\widehat{HERF}$	0.239***	0.236***	0.235***	0.231***	0.241***	0.240***	0.237***	0.234***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
$lnFUEL$	0.120***	0.055***	0.104***	0.037***	0.126***	0.055***	0.110***	0.037***
	(0.012)	(0.013)	(0.013)	(0.015)	(0.011)	(0.013)	(0.012)	(0.014)
$lnCOST$	0.173***	0.096***	0.148***	0.067	0.176**	0.093*	0.150***	0.063
	(0.061)	(0.055)	(0.057)	(0.052)	(0.061)	(0.055)	(0.057)	(0.051)
Observations	154390	154390	153689	153689	154390	154390	153689	153689

Table 3.7: Estimates with Capacity Utilization Control: Gini

All regressions include carrier-route-specific dummies and quarter dummies. IV estimates indicate that UTIL and UTILdt were instrumented using utilization at the route-quarter level. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10 per cent, 5 per cent, or 1 per cent significance level, respectively.

$Gini^{lodd}$								
	OLS				IV			
YGAP	1.235***	1.375***			1.233***	1.301***		
	(0.106)	(0.110)			(0.105)	(0.108)		
-UR			1.752***	1.923***			1.751***	1.833***
			(0.155)	(0.160)			(0.155)	(0.158)
$UTIL$	-0.305***		-0.303***		-0.299***		-0.298***	
	(0.036)		(0.036)		(0.043)		(0.043)	
$UTIL_{dt}$		-0.475***		-0.469***		-0.321***		-0.320***
		(0.040)		(0.040)		(0.047)		(0.047)
$\ln \widehat{HERF}$	0.072***	0.068***	0.069***	0.065***	0.072***	0.071***	0.069***	0.068***
	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)	(0.023)
$\ln FUEL$	-0.022***	-0.039***	-0.033***	-0.051***	-0.023***	-0.040***	-0.033***	-0.051***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
$\ln COST$	0.460***	0.434***	0.445***	0.417***	0.460***	0.439***	0.445***	0.423***
	(0.030)	(0.029)	(0.030)	(0.028)	(0.031)	(0.029)	(0.030)	(0.028)
Observations	156021	156021	155314	155314	156021	156021	155314	155314

Bibliography

Adler, G. and C. Tovar, "Foreign Exchange Intervention: A Shield Against Appreciation Winds?," *IMF working paper*, 2011.

Aguair, M. and M. Amador, "Growth in the shadow of expropriation," *The Quarterly Journal of Economics*, 2011, *126*, 651–697.

Aguirre, I., S. Cowan, and J. Vickers, "Monopoly Price Discrimination and Demand Curvature," *American Economic Review*, 2010, *100*, 1601–1615.

Aizcorbe, A. and A. Shapiro, "Implications of Consumer Heterogeneity on Price Measures for Technology Goods," *BEA Working Paper*, 2010.

Aizenman, J. and J. Lee, "International Reserves: Precautionary Versus Mercantilistic Views, Theory and Evidence," *Open Economies Review*, 2007, *18*, 191–214.

_ **and R. Glick**, "Sterilization, Monetary Policy, and Global Financial Integration," *Review of International Economics*, 2009, *17* (4), 777–801.

Bacchetta, P., K. Benhima, and Y. Kalantzis, "Capital Controls with International Reserve Accumulation: Can this Be Optimal?," *American Economic Journal: Macroeconomics*, 2013, *5*, 229–262.

Bacchiocchi, E. and L. Fanelli, "Identification in Structural Vector Autoregressive Models with Structural Changes," *Working Paper*, 2012.

_ **, E. Castelnuovo, and L. Fanelli**, "Gimme a Break! Identification and Estimation of Macroeconomic Effects of Monetary Policy Shocks in the U.S.," *Working Paper*, 2014.

Baye, M., J. Morgan, and P. Scholten, "Price Dispersion in the Small and the Large: Evidence from an Internet Price Comparison Site," *The Journal of Industrial Economics*, 2004, *52*, 463–496.

Bayoumi, T., J. Gagnon, and C. Saborowski, "Official Financial Flows, Capital Mobility, and Global Imbalances," *Journal of International Money and Finance*, 2015, *52*, 146174.

Benigno, G. and L. Fornaro, "Reserve Accumulation, Growth and Financial Crises," *Working Paper*, 2012.

Bernanke, B., "Nonmonetary effects of the financial crisis in the propagation of the great depression," *American Economic Review*, 1983, *73*, 257–276.

_ , "The global saving glut and the U.S. current account deficit," *Speech to the Virginia Association of Economics*, 2005.

Bester, H., "The role of collateral in credit markets with imperfect information," *European Economic Review*, 1987, *31*, 887–899.

Bils, M., "Pricing in a Customer Market," *Quarterly Journal of Economics*, 1989, *104*, 699–718.

Borenstein, S. and N. Rose, "Competition and Price Dispersion in the U.S. Airline Industry," *The Journal of Political Economy*, 1994, *102*, 653–683.

Burdett, K. and K. Judd, "Equilibrium Price Dispersion," *Econometrica*, 1983, *51*, 955–969.

Busse, M. and M. Rysman, "Competition and Price Discrimination in Yellow Pages Advertising," *The RAND Journal of Economics*, 2005, *36*, 378–390.

Carroll, C. and O. Jeanne, "A tractable model of precautionary reserves or net foreign assets," *NBER Working Paper*, 2009.

Carter, C.K. and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, 1994, *81*, 541–553.

Caves, D., R. Christensen, and M. Tretheway, "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ," *The RAND Journal of Economics*, 1984, *15*, 471–489.

Chamberlain, G., "Panel Data," *Handbook of Econometrics*, 1984, *2*, 1247–1318.

Chang, C., Z. Liu, and M. Spiegel, "Monetary policy in a DSGE model with Chinese Characteristics," *Working Paper*, 2013.

Cheung, Y. and H. Ito, "A Cross-Country Empirical Analysis of International Reserves," *International Economic Journal*, 2009, *23*, 447–481.

Chevalier, J., A. Kashyap, and P. Rossi, "Why don't prices rise during periods of peak demand? Evidence from Scanner Data," *American Economic Review*, 2003, *93*, 15–37.

__ **and D. Scharfstein**, "Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence," *American Economic Review*, 1996, *86*, 703–725.

Contessi, S. and J. Francis, "U.S. Commercial bank lending through 2008:Q4: New evidence from gross credit flows," *Economic Inquiry*, 2010, *51*, 428–444.

Craig, B. and J. Haubrich, "Gross Loan Flows," *Journal of Money, Credit and Banking*, 2013, *45*, 401–421.

Dana, J., "Equilibrium Price Dispersion Under Demand Uncertainty: The Roles of Costly Capacity and Market Structure," *The RAND Journal of Economics*, 1999, *30*, 632–660.

Davis, S. and J. Haltiwanger, "Gross job creation, gross job destruction, and employment reallocation," *Quarterly Journal of Economics*, 1992, *3*, 819–863.

Dell'Arriccia, G. and P. Garibaldi, "Bank lending and interest rate changes in a dynamic matching model," *International Monetary Fund Working Paper*, 1998.

__ **and** __ , "Gross Credit Flows," *Review of Economics and Statistics*, 2005, *72*, 665–685.

Disyatat, P. and G. Galati, "The effectiveness of foreign exchange intervention in emerging market countries: Evidence from the Czech koruna," *Journal of International Money and Finance*, 2007, *26*, 383–402.

Dominguez, K., "Foreign reserve management during the global financial crisis," *Journal of International Money and Finance*, 2012, *31*, 2017–2037.

_, Y. Hashimoto, and T. Ito, "International reserves and the global financial crisis," *Journal of International Economics*, 2012.

Duarte, M., "Why don't macroeconomic quantities respond to exchange rate variability?," *Journal of Monetary Economics*, 2003, *50*, 889–913.

Eden, B., "Marginal Cost Pricing When Spot Markets are Complete," *Journal of Political Economy*, 1990, *98*, 1293–1306.

Engel, C. and J. Rogers, "The U.S. Current account deficit and the expected share of world output," *Journal of Monetary Economics*, 2006, *53*, 1063–1093.

Farhi, E. Caballero R. and P. Gourinchas, "An Equilibrium model of Global Imbalances and Low interest rates," *American Economic Review*, 2008, *98*, 358–393.

Faust, J., E. Swanson, and J. Wright, "Identifying VARS based on high frequency futures data," *Journal of Monetary Economics*, 2004, *51*, 1107–1131.

Folkerts-Landau, D. Dooley M. and P. Garber, "An essay on the revived Bretton Woods System," *Working Paper No. 9971 National Bureau of Economic Research*, 2003.

Fratzscher, M., "On the long-term effectiveness of exchange rate communication and interventions," *Journal of International Money and Finance*, 2006, *25*, 146–167.

Frenkel, J. and S. Wei, "Assessing China's exchange rate regime," *Economic Policy*, 2007, *22*, 575–627.

Gabaix, X., "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 2011, *79*, 733–772.

Gaggero, A. and C. Piga, "Airline Market Power and Intertemporal Price Dispersion," *Journal of Industrial Economics*, 2011, *59*, 552–557.

Gerardi, K. and A. Shapiro, "Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry," *Journal of Political Economy*, 2009, *117*, 1–37.

Gillen, D., T. Oum, and M. Tretheway, "Airline Cost Structure and Policy Implications: A Multi-Product Approach for Canadian Airlines," *Journal of Transport Economics and Policy*, 1990, *24*, 9–34.

Goldberg, P. and F. Verboven, "The Evolution of Price Dispersion in the European Car Market," *Review of Economic Studies*, 2001, *68*, 811–848.

Goolsbee, A. and C. Syverson, "How do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines," *The Quarterly Journal of Economics*, 2008, *123*, 1611–1633.

Greenwald, B., J. Stiglitz, and A. Weiss, "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations," *American Economic Review Papers and Proceedings*, 1984, *74*, 194–199.

Hendel, I. and A. Nevo, "Intertemporal Price Discrimination in Storable Goods Markets," *NBER Working Paper No. 16988*, 2011.

Herrera, A., M. Kolar, and R. Minetti, "Credit reallocation," *Journal of Monetary Economics*, 2011, *58*, 551–563.

Holmes, T., "The Effect of Third Degree Price Discrimination in Oligopoly," *American Economic Review*, 1989, *79*, 244–250.

Inoue, A. and L. Kilian, "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics*, 2013, *177*, 1–13.

Ito, H. and D. Lee, "Low Cost Carrier Growth in the U.S. Airline Industry: Past, Present and Future," *Brown University Department of Economics Paper No. 2003-12*, 2003.

Jeanne, O., "International Reserves in Emerging Market Countries: Too Much of a Good Thing," *Brookings Papers on Economic Activity*, Dec 2007, pp. 1–55.

_, "Capital Account Policies and the Real Exchange Rate," *PIIE Working Paper, WP 12-14*, 2012.

_ **and R. Ranciere**, "The Optimal Level of International Reserves For Emerging Market Countries: A New Formula and Some Applications," *The Economic Journal*, 2011, *121*, 905–930.

Jikang, Z. and L. Yuanyuan, "Institutional and Structural Problems of China's Foreign Exchange Market and the RMB's Role in East Asia," *HWWI Research Paper*, 2006.

Kearns, J. and R. Rigobon, "Identifying the efficacy of central bank interventions: evidence from Australia and Japan," *Journal of International Economics*, 2005, *66*, 31–48.

Klemperer, P., "Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade," *Review of Economic Studies*, 1995, *62*, 515–539.

Korinek, A. and L. Serven, "Undervaluation through Foreign Reserve Accumulation: Static Losses, Dynamic Gains," *World Bank Policy Research Working Paper No. 5250*, 2010.

Kumhof, M, "On the theory of sterilized foreign exchange intervention," *Journal of International Economics*, 2010, *34*, 1403–1420.

Leeper, E. and T. Zha, "Modest Policy Interventions," *Journal of Monetary Economics*, 2003, *50*, 1673–1700.

Leslie, P., "Price Discrimination in the Broadway Theater," *The RAND Journal of Economics*, 2004, *35*, 520–541.

Levy-Yeyati, E. and F. Sturzenegger, "Classifying exchange rate regimes: Deeds vs. Words," *European Economic Review*, 2005, *49*, 1603–1635.

Lucas, R., "Expectations and the neutrality of money," *Journal of Economic Theory*, 1972, *4*, 103–124.

MacDonald, J., "Demand, Information and Competition: Why Do Food Prices Fall at Seasonal Demand Peaks?," *Journal of Industrial Economics*, 2000, *48*, 27–45.

Martin, C., "Identifying the Effect of Chinese Treasury Purchases Using High Frequency Data," *Working Paper*, 2013.

Maskin, E. and J. Riley, "Monopoly with Incomplete Information," *The RAND Journal of Economics*, 1984, *15*, 171–196.

Miller, N. and M. Osborne, "Competition among Spatially Differentiated Firms. An Empirical Model with an Application to Cement," *EAG Working Paper 10-2*, 2010.

Mussa, M. and S. Rosen, "Monopoly and product quality," *Journal of Economic Theory*, 1978, *18*, 301–317.

Nevo, A. and K. Hatzitaskos, "Why Does the Average Price Paid Fall During High Demand Periods?," *NBER Working Paper No. 11572*, 2006.

Orlov, E., "How Does the Internet Influence Price Dispersion? Evidence from the Airline Industry," *Journal of Industrial Economics*, 2011, *59*, 21–37.

Ouyang, A., R. Rajan, and T. Willett, "China as a reserve sink: The evidence from offset and sterilization coefficients," *Journal of International Money and Finance*, 2010, *29* (5), 951–972.

Pennacchi, G., "Deposit insurance, bank regulation, and financial system risks," *Journal of Monetary Economics*, 2006, *53*, 1–30.

Petersen, M. and R. Rajan, "Does distance still matter? The information revolution in small business lending," *Journal of Finance*, 2002, *57*, 2533–2570.

Prasad, E. and S.J. Wei, "The Chinese Approach to Capital Inflows: Patterns and Possible Explanations," *NBER Working Paper*, 2005.

Preston, H., "The Banking Act of 1933," *American Economic Review*, 1933, *23*, 585–607.

Puller, S., A. Sengupta, and S. Wiggins, "Testing Theories of Scarcity Pricing in the Airline Industry," *NBER Working Paper 15555*, 2009.

Quadrini, V. Mendoza E. and J. Rios-Rull, "Financial Integration, Financial Deepness and Global Imbalances," *CEPR Discussion Papers 6149*, 2007.

Rigobon, R., "Identification Through Heteroskedasticity," *The Review of Economics and Statistics*, 2003, *85*, 777–792.

Rodrik, D., "The Real Exchange Rate and Economic Growth," *Brookings Papers on Economic Activity*, 2008, pp. 365–412.

Romer, C., "The Cyclical Behaviour of Individual Production Series, 1889-1984," *The Quarterly Journal of Economics*, 1991, *106*, 1–31.

Rotenmberg, J. and G. Saloner, "A Super-game Theoretic Model of Business Cycles and Price Wars During Booms," *American Economic Review*, 1986, *76*, 390–407.

Salop, S. and J. Stiglitz, "The Theory of Sales: A simple Model of Equilibrium Price Dispersion with Identical Agents," *American Economic Review*, 1982, *72*, 1121–1130.

Sarno, L. and M Taylor, "Official Intervention in the Foreign Exchange Market: Is It Effective, and, If So, How Does It Work?," *Journal of Economic Literature*, September 2001, *39*, 838–868.

Sheng, L., "Did China diversify its foreign reserves?," *Journal of Applied Econometrics*, 2013, *28*, 102–125.

Shepard, A., "Price Discrimination and Retail Configuration," *Journal of Political Economy*, 1991, *99*, 30–53.

Sorensen, A., "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy*, 2000, *108*, 833–850.

Stavins, J., "Price Discrimination in the Airline Market: The Effect of Market Concentration," *Review of Economics and Statistics*, 2001, *83*, 200–202.

Stiglitz, J., "Price Rigidities and Market Structure," *American Economic Review*, 1984, *74*, 350–355.

Tirole, J., "The Theory of Industrial Organization," *MIT Press, Cambridge*, 1988.

Uhlig, H., "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 2005, *52*, 381–419.

Vegh, C., *Open Economy Macroeconomics in Developing Countries*, The MIT Press, 2013.

Warner, E. and R. Barsky, "The Timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays," *Quarterly Journal of Economics*, 1995, *110*, 321–352.

Zhang, C., "Sterilization in China: Effectiveness and Cost," *Working Paper*, 2010.

CHAPTER 4

Curriculum Vitae

Marco C. Cornia received a B.A. degree in Economics and Finance from the University of Bologna, Italy, in 2007. In 2008 he received a M.Sc. in Economics from the London School of Economics, United Kingdom. He enrolled in the Ph.D. program in Economics at the Johns Hopkins University in 2009.